

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number  
WO 01/27284 A2

(51) International Patent Classification<sup>7</sup>: C12N 15/52,  
15/53, 15/54, 15/61, 15/62, 9/04, 9/10, 9/90, C12P 19/62

(21) International Application Number: PCT/US00/27433

(22) International Filing Date: 5 October 2000 (05.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/158,305 8 October 1999 (08.10.1999) US  
60/190,024 17 March 2000 (17.03.2000) US

(71) Applicant: KOSAN BIOSCIENCES, INC. [US/US];  
3832 Bay Center Place, Hayward, CA 94545 (US).

(72) Inventors: MCDANIEL, Robert; Palo Alto, CA (US).  
VOLCHEGURKSY, Yanina; Emeryville, CA (US).

(74) Agents: CHEN, Peng et al.; Morrison & Foerster LLP,  
12636 High Bluff Drive, Suite 300, San Diego, CA 92130-  
2071 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

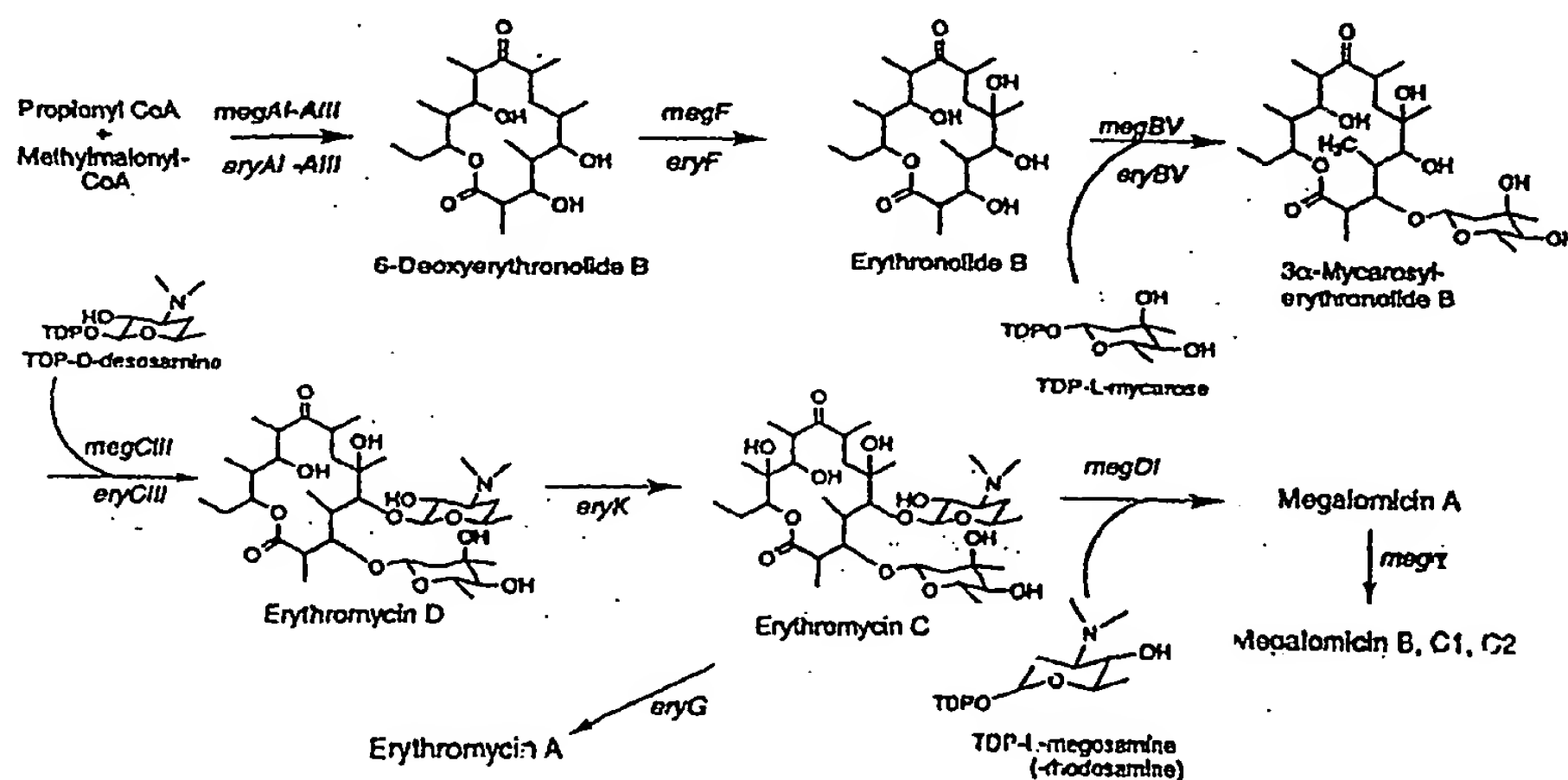
(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: RECOMBINANT MEGALOMICIN BIOSYNTHETIC GENES AND USES THEREOF



(57) Abstract: Recombinant nucleic acid, e.g. DNA compounds that encode all or a portion of the megalomicin polyketide synthase and modification enzymes are used to express recombinant polyketide synthase genes in host cells for the production of megalomicin, megalomicin derivatives, and other polyketides that are useful as antibiotics, motilides, and antiparasitics.

WO 01/27284 A2



### Title

## **Recombinant Megalomycin Biosynthetic Genes And Uses Thereof**

### Cross-Reference to Priority Application

5        This application claims priority to provisional U.S. patent application  
Serial No. 60/158,305, filed 8 October 1999, and provisional U.S. patent  
application Serial No. 60/190,024, filed 17 March 2000 under 35 U.S.C. § 119(e).  
The content of the above referenced applications is incorporated herein by  
reference in its entirety.

10

### Field of the Invention

      The present invention provides recombinant methods and materials for  
producing polyketides by recombinant DNA technology. The invention relates to  
15    the fields of agriculture, animal husbandry, chemistry, medicinal chemistry,  
medicine, molecular biology, pharmacology, and veterinary technology.

### Background of the Invention

      Polyketides represent a large family of diverse compounds synthesized  
20    from 2-carbon units through a series of condensations and subsequent  
modifications. Polyketides occur in many types of organisms, including fungi and  
mycelial bacteria, in particular, the actinomycetes. There are a wide variety of  
polyketide structures, and the class of polyketides encompasses numerous  
compounds with diverse activities. Erythromycin, FK-506, FK-520, megalomicin,  
25    narbomycin, oleandomycin, picromycin, rapamycin, spinocyn, and tylosin are  
examples of such compounds. Given the difficulty in producing polyketide  
compounds by traditional chemical methodology, and the typically low production  
of polyketides in wild-type cells, there has been considerable interest in finding  
improved or alternate means to produce polyketide compounds. See PCT  
30    publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358;  
and WO 98/27203; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837;  
5,149,639; 5,672,491; and 5,712,146; Fu *et al.*, 1994, *Biochemistry* 33: 9321-  
9326; McDaniel *et al.*, 1993, *Science* 262: 1546-1550; and Rohr, 1995, *Angew.*

*Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by polyketide synthase (PKS) enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKS enzymes are encoded by PKS genes that usually consist of three or more open reading frames (ORFs). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

Modular PKSs are responsible for producing a large number of 12-, 14-, and 16-membered macrolide antibiotics including erythromycin, megalomicin, methymycin, narbomycin, oleandomycin, picromycin, and tylosin. Each ORF of a modular PKS can comprise one, two, or more "modules" of ketosynthase activity, each module of which consists of at least two (if a loading module) and more typically three (for the simplest extender module) or more enzymatic activities or "domains." These large multifunctional enzymes (>300,000 kDa) catalyze the biosynthesis of polyketide macrolactones through multistep pathways involving decarboxylative condensations between acyl thioesters followed by cycles of varying  $\beta$ -carbon processing activities (see O'Hagan, D. *The polyketide metabolites*; E. Horwood: New York, 1991, incorporated herein by reference).

During the past half decade, the study of modular PKS function and specificity has been greatly facilitated by the plasmid-based *Streptomyces coelicolor* expression system developed with the 6-deoxyerythronolide B (6-dEB) synthase (DEBS) genes (see Kao *et al.*, 1994, *Science*, 265: 509-512, McDaniel *et al.*, 1993, *Science* 262: 1546-1557, and U.S. Patent Nos. 5,672,491 and 5,712,146, each of which is incorporated herein by reference). The advantages to this plasmid-based genetic system for DEBS are that it overcomes the tedious and limited techniques for manipulating the natural DEBS host organism, *Saccharopolyspora erythraea*, allows more facile construction of recombinant PKSs, and reduces the complexity of PKS analysis by providing a "clean" host background. This system also expedited construction of the first combinatorial

modular polyketide library in *Streptomyces* (see PCT publication No. WO 98/49315, incorporated herein by reference).

The ability to control aspects of polyketide biosynthesis, such as monomer selection and degree of  $\beta$ -carbon processing, by genetic manipulation of PKSs has stimulated great interest in the combinatorial engineering of novel antibiotics (see Hutchinson, 1998, *Curr. Opin. Microbiol.* 1: 319-329; Carreras and Santi, 1998, *Curr. Opin. Biotech.* 9: 403-411; and U.S. Patent Nos. 5,712,146 and 5,672,491, each of which is incorporated herein by reference). This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters.

Megalomicin is a macrolide antibiotic produced by *Micromonospora megalomicea*, a member of the Actinomycetales family of soil bacteria that produces many types of biologically active compounds. Megalomicin is a glycoside of erythromycin A, a widely used antibacterial drug with little or no antimalarial activity. Megalomicin has antibacterial properties similar to those of erythromycin, and in 1998, it was discovered also to have potent antiparasitic activity and low toxicity. The antiparasitic activity may be related to the effect megalomicin has on protein trafficking in eukaryotes, where it appears to inhibit vesicular transport between the medial and trans-Golgi, resulting in under-sialylation of proteins. Hence, megalomicin offers an exciting opportunity to develop a new class of antiparasitic drugs with a different mechanism of action than the drugs currently in use and, therefore, possibly active against drug-resistant forms of *Plasmodium falciparum*.

The number and diversity of megalomicin derivatives have been limited due to the inability to manipulate the PKS genes, which have not previously been available in recombinant form. Genetic systems that allow rapid engineering of the megalomicin biosynthetic genes would be valuable for creating novel compounds for pharmaceutical, agricultural, and veterinary applications. The production of

such compounds could be more readily accomplished if the heterologous expression of the megalomicin biosynthetic genes in *Streptomyces coelicolor* and *S. lividans* and other host cells were possible. The present invention meets these and other needs.

5

### Summary of the Invention

The present invention provides recombinant methods and materials for expressing PKS enzymes and polyketide modification enzymes derived in whole and in part from the megalomicin biosynthetic genes in recombinant host cells.

10 The invention also provides the polyketides produced by such PKS enzymes. The invention provides in recombinant form all of the genes for the proteins that constitute the complete PKS that ultimately results, in *Micromonospora megalomicea*, in the production of megalomicin. Thus, in one embodiment, the invention is directed to recombinant materials comprising nucleic acids with  
15 nucleotide sequences encoding at least one domain, module, or protein encoded by a megalomicin PKS gene. In one preferred embodiment of the invention, the DNA compounds of the invention comprise a coding sequence for at least one and preferably two or more of the domains of the loading module and extender modules 1 through 6, inclusive, of the megalomicin PKS.

20 In one embodiment, the invention provides a recombinant expression vector that comprises a heterologous promoter positioned to drive expression of one or more of the megalomicin biosynthetic genes. In a preferred embodiment, the promoter is derived from another PKS gene. In a related embodiment, the invention provides recombinant host cells comprising one or more expression  
25 vectors that produce(s) megalomicin or a megalomicin derivative or precursor. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In another embodiment, the invention provides a recombinant expression vector that comprises a promoter positioned to drive expression of a hybrid PKS comprising all or part of the megalomicin PKS and at least a part of a second PKS.  
30 In a related embodiment, the invention provides recombinant host cells comprising the vector that produces the hybrid PKS and its corresponding polyketide. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In a related embodiment, the invention provides recombinant materials for the production of libraries of polyketides wherein the polyketide members of the library are synthesized by hybrid PKS enzymes of the invention. The resulting polyketides can be further modified to convert them to other useful compounds, such as antibiotics, motilides, and antiparasitics, typically through hydroxylation and/or glycosylation. Modified macrolides provided by the invention that are useful intermediates in the preparation of antiparasitics are of particular benefit.

In another related embodiment, the invention provides a method to prepare a nucleic acid that encodes a modified PKS, which method comprises using the megalomicin PKS encoding sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement. The thus modified megalomicin PKS encoding nucleotide sequence can then be expressed in a suitable host cell and the cell employed to produce a polyketide different from that produced by the megalomicin PKS. In addition, portions of the megalomicin PKS coding sequence can be inserted into other PKS coding sequences to modify the products thereof.

In another related embodiment, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression vector for the production of a modular PKS derived in whole or in part from the megalomicin PKS. Thus, at least a portion of the modular PKS is identical to that found in the PKS that produces megalomicin and is identifiable as such. The derived portion can be prepared synthetically or directly from DNA derived from organisms that produce megalomicin. In addition, the invention provides methods to screen the resulting polyketide and antibiotic libraries.

The invention also provides novel polyketides, motilides, antibiotics, antiparasitics and other useful compounds derived therefrom. The compounds of the invention can also be used in the manufacture of another compound. In a preferred embodiment, the compounds of the invention are formulated in a mixture or solution for administration to an animal or human.

In a specific embodiment, the invention provides an isolated nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The isolated

nucleic acid fragment can be a DNA or a RNA. Preferably, the isolated nucleic acid fragment is a recombinant DNA compound.

The isolated nucleic acid fragment can comprise a single, multiple or all the open reading frame(s) (ORF) of the megalomicin PKS or a megalomicin modification enzyme. Exemplary ORFs of megalomicin PKS include the ORFs of the *megAI*, *megAII* and *megAIII* genes. The isolated nucleic acid fragment can also encode a single, multiple, or all of the domains of the megalomicin PKS. Exemplary domains of the megalomicin PKS include a TE domain, a KS domain, an AT domain, an ACP domain, a KR domain, a DH domain and an ER domain. In a preferred embodiment, the nucleic acid fragment encodes a module of the megalomicin PKS. In another preferred embodiment, the nucleic acid fragment encodes the loading module, a thioesterase domain, and all six extender modules of the megalomicin PKS.

Megalomicin modification enzymes include those enzymes involved in the conversion of 6-dEB into a megalomicin such as the enzymes encoded by the *megF*, *megBV*, *megCIII*, *megK*, *megDI* and *megG* (renamed *megY*) genes. Megalomicin modification enzymes also include those enzymes involved in the biosynthesis of mycarose, megosamine or desosamine, which are used as biosynthetic intermediates in the biosynthesis of various megalomicin species and other related polyketides. The enzymes that are involved in biosynthesis of mycarose, megosamine or desosamine are described in Figures 5 and 10.

In a preferred embodiment, the invention provides an isolated nucleic acid fragment which hybridizes to a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1, under low, medium or high stringency. More preferably, the nucleic acid fragment comprises, consists or consists essentially of a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1.

In another specific embodiment, the invention provides a substantially purified polypeptide, which is encoded by a nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The polypeptide can comprise a single domain, multiple domains or a full-length megalomicin PKS or megalomicin modification enzyme. Functional fragments, analogs or derivatives of the megalomicin PKS or megalomicin modification enzyme polypeptides are



also provided. Preferably, such fragments, analogs or derivatives can be recognized by an antibody raised against a megalomicin PKS or megalomicin modification enzyme. Also preferably, such fragments, analogs or derivatives comprise an amino acid sequence that has at least 60% identity, more preferably at least 90% identity, to their wild type counterparts.

In still another specific embodiment, the invention provides an antibody, or a fragment or derivative thereof, which immuno-specifically binds to a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The antibody can be a monoclonal or polyclonal antibody or an antibody fragment. Preferably, the antibody is a monoclonal antibody.

In yet another specific embodiment, the invention provides a recombinant DNA expression vector comprising the recombinant DNA compound encoding at least a domain of the megalomicin PKS or a megalomicin modification enzyme, wherein said domain is operably linked to a promoter. Preferably, the recombinant DNA expression vector further comprises an origin of replication or a segment of DNA that enables chromosomal integration.

In yet another specific embodiment, the invention provides a recombinant host cell comprising the above-described recombinant DNA expression vector encoding at least a domain of megalomicin PKS or the megalomicin modification enzyme. The recombinant host cells can be any suitable host cells including animal, mammalian, plant, fungal, yeast, and bacterial cells. Preferably, the recombinant host cells are *Streptomyces* cells, such as *Streptomyces lividans* and *S. coelicolor* cells, or *Saccharopolyspora* cells, such as *Saccharopolyspora erythraea* cells. Also preferably, the recombinant host cells do not produce megalomicin in their untransformed, non-recombinant state.

When the recombinant host cell contains nucleic acid encoding more than one megalomicin PKS or megalomicin modification enzyme, or domains thereof, such nucleic acid material can be located at a single genetic locus, e.g., on a single plasmid or at a single chromosomal locus, or at different genetic loci, e.g., on separate plasmids and/or chromosomal loci. In one example, the invention provides a recombinant host cell, which comprises at least two separate autonomously replicating recombinant DNA expression vectors, and each of said vectors comprises a recombinant DNA compound encoding a megalomicin PKS

domain or a megalomicin modification enzyme operably linked to a promoter. In another example, the invention provides a recombinant host cell, which comprises at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant DNA compound encoding a megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter. Preferably, the autonomously replicating recombinant DNA expression vector and/or the modified chromosome further comprises distinct selectable markers.

10 In a preferred embodiment, the cell comprises three different vectors, one of which is integrated into the chromosome and two of which are autonomously replicating, and each of the vectors comprises a *meg* PKS gene. Optionally, one or more of the *meg* PKS genes contains one or more domain alterations, such as a deletion or substitution of a *meg* PKS domain with a domain from another PKS.

15 In yet another specific embodiment, the invention provides a hybrid PKS, which is produced from a recombinant gene that comprises at least a portion of a megalomicin PKS gene and at least a portion of a second PKS gene for a polyketide other than megalomicin. For example, and without limitation, the second PKS gene can be a narbonolide PKS gene, an oleandolide PKS gene, or a rapamycin PKS gene. In one embodiment, the hybrid PKS is composed of a loading module and six extender modules, wherein at least one domain of any one of extender modules 1 through 6, inclusive, is a domain of an extender module of megalomicin PKS. In another preferred embodiment, the hybrid PKS comprises a megalomicin PKS that has a non-functional KS domain in module 1.

25 In yet another specific embodiment, the invention provides a method of producing a polyketide, which method comprises growing the recombinant host cell comprising a recombinant DNA expression vector encoding at least a domain of the megalomicin PKS or a megalomicin modification enzyme under conditions whereby the megalomicin PKS domain or the megalomicin modification enzyme comprised by the recombinant expression vector is produced and the polyketide is synthesized by the cell, and recovering the synthesized polyketide. Preferably, the recombinant host cell comprises a recombinant expression vector that encodes at least a portion of a *megAI*, *megAII*, or *megAIII* gene.



These and other embodiments of the invention are described in more detail in the following description, the examples, and claims set forth below.

### Brief Description of the Figures

5           Figure 1 shows restriction site and function maps of the insert DNA in cosmids pKOS079-138B, pKOS079-93D, pKOS079-93A, and pKOS079-124B of the invention. Various restriction sites (*Xho*I, *Bgl*II, *Nsi*I) are also shown. The location of the megalomicin biosynthetic genes is shown below the solid lines indicating the cosmid inserts. The genes are shown as arrows pointing in the  
10   direction of transcription. The approximate size (in kilobase (kb) pairs) of the gene cluster is indicated in 5000 bp (i.e., 5K, 10K, and the like.) increments on a solid bar beneath the arrows indicating the genes.

          Figure 2 shows a more detailed map of the megalomicin biosynthetic gene cluster. The various open reading frames are shown as arrows pointing in the  
15   direction of transcription. A line indicates the size in base pairs (in 1000 bp increments) of the gene cluster. The various domains of the megalomicin PKS are also shown. Other genes of the megalomicin biosynthetic gene cluster not shown in this Figure are located in the insert DNA of cosmids pKOS0138B and pKOS0124B.

20           Figure 3 shows the structures of the megalomicins, azithromycin and erythromycin A.

          Figure 4 shows the modules and domains of DEBS and the megalomicin PKS.

          Figure 5 shows the compounds and reactions in the erythromycin  
25   biosynthetic pathway and also for megalomicin biosynthesis. Genes that produce the various enzymes that catalyze each of the steps in the biosynthetic pathway are indicated.

          Figure 6 shows the biosynthetic pathway for the formation of desosamine, rhodosamine, and mycarose, as well as the genes that produce the various enzymes  
30   that catalyze each of the steps in the biosynthetic pathway.

          Figure 7 depicts nucleotide and amino acid sequence of *Micromonospora megalomicea* megalomicin biosynthetic genes (GenBank Accession No. AF263245, incorporated herein by reference).

Figure 8 depicts the biosynthesis of the erythromycins and megalomicins and the enzymes that mediate the biosynthesis of each.

Figure 9 depicts the cloned megalomicin biosynthetic gene cluster and certain cosmids of the invention that comprise portions of the cluster.

5 Figure 10 depicts the biosynthesis of megosamine, mycarose, and desosamine.

### Detailed Description of the Invention

The present invention provides useful compounds and methods for  
10 producing polyketides in recombinant host cells. As used herein, the term recombinant refers to a compound or composition produced by human intervention. The invention provides recombinant DNA compounds encoding all or a portion of the megalomicin biosynthetic genes. The invention provides recombinant expression vectors useful in producing the megalomicin PKS and  
15 hybrid PKSs composed of a portion of the megalomicin PKS in recombinant host cells. The invention also provides the polyketides produced by the recombinant PKS and polyketide modification enzymes.

To appreciate the many and diverse benefits and applications of the invention, the description of the invention below is organized as follows. In  
20 Section I, common definitions used throughout this application are provided. In Section II, structural and functional characteristics of megalomicin are described. In Section III, the recombinant megalomicin biosynthetic genes and other recombinant nucleic acids provided by the invention are described. In Section IV, polypeptides and proteins encoded by the megalomicin biosynthetic genes and  
25 antibodies that specifically bind to such polypeptides and proteins provided by the invention are described. In Section V, methods for heterologous expression of the megalomicin biosynthetic genes provided by the invention are described. In Section VI, the hybrid PKS genes provided by the invention are described. In Section VII, host cells containing multiple megalomicin biosynthetic genes and  
30 nucleic acid fragments on separate express vectors provided by the invention are described. In Section VIII, the polyketide compounds provided by the invention and pharmaceutical compositions of those compounds are described. The detailed description is followed by working examples illustrating the invention.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. All patents, applications, published applications and other publications and sequences from GenBank and other data  
5 bases referred to herein are incorporated by reference in their entirety.

### Section I. Definitions

As used herein, domain refers to a portion of a molecule, e.g., proteins or nucleic acids, that is structurally and/or functionally distinct from another portion  
10 of the molecule.

As used herein, antibody includes antibody fragments, such as Fab fragments, which are composed of a light chain and the variable region of a heavy chain.

As used herein, biological activity refers to the *in vivo* activities of a  
15 compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in in vitro systems designed to test or use such activities.

20 As used herein, a combination refers to any association between two or among more items.

As used herein, a composition refers to any mixture. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

25 As used herein, derivative or analog of a molecule refers to a portion derived from or a modified version of the molecule.

As used herein, operably linked, operatively linked or operationally associated refers to the functional relationship of DNA with regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional  
30 and translational stop sites, and other signal sequences. For example, operative linkage of DNA to a promoter refers to the physical and functional relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes,

binds to and transcribes the DNA. To optimize expression and/or *in vitro* transcription, it may be helpful to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potentially inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak, *J. Biol. Chem.*, 266:19867-19870 (1991)) can be inserted immediately 5' of the start codon and may enhance expression. The desirability of (or need for) such modification may be empirically determined.

10 As used herein, pharmaceutically acceptable salts, esters or other derivatives of the conjugates include any salts, esters or derivatives that may be readily prepared by those of skill in this art using known methods for such derivatization and that produce compounds that may be administered to animals or humans without substantial toxic effects and that either are pharmaceutically  
15 active or are prodrugs.

As used herein, a promoter region or promoter element refers to a segment of DNA or RNA that controls transcription of the DNA or RNA to which it is operatively linked. The promoter region includes specific sequences that are sufficient for RNA polymerase recognition, binding and transcription initiation.  
20 This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of RNA polymerase. These sequences may be *cis* acting or may be responsive to *trans* acting factors. Promoters, depending upon the nature of the regulation, may be constitutive or regulated.

25 As used herein: stringency of hybridization in determining percentage mismatch is as follows: (1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C; (2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C; and (3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C. Equivalent stringencies may be achieved using alternative buffers, salts and temperatures.

The term substantially identical or homologous or similar varies with the context as understood by those skilled in the relevant art and generally means at least 70%, preferably means at least 80%, more preferably at least 90%, and most preferably at least 95% identity.

5           As used herein, substantially identical to a product means sufficiently similar so that the property of interest is sufficiently unchanged so that the substantially identical product can be used in place of the product.

          As used herein, isolated means that a substance is either present in a preparation at a concentration higher than that substance is found in nature or in its  
10   naturally occurring state or that the substance is present in a preparation that contains other materials with which the substance is not associated with in nature. As an example of the latter, an isolated meg PKS protein includes a meg PKS protein expressed in a *Streptomyces coelicolor* or *S. lividans* host cell.

          As used herein, substantially pure means sufficiently homogeneous to  
15   appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and  
20   biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous DNA into cells for either expression or replication thereof. Selection and use of such vehicles are well known within the skill of the artisan. An expression vector includes vectors capable of expressing DNAs that are operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

## Section II. Megalomicins

The megalomicins were discovered in 1969 at Schering Corp. as antibacterial agents produced by *Micromonospora megalomicea* (see Weinstein *et al.*, 1969, *J. Antibiotics* 22: 253-258, and U.S. Patent No. 3,632,750, both of which are incorporated herein by reference). Although the initial structural assignment was in error, a thorough reassessment of NMR data coupled with an X-ray crystal structure of a megalomicin A derivative (see Nakagawa and Omura, "Structure and Stereochemistry of Macrolides" in *Macrolide Antibiotics* (S. Omura, ed.), Academic Press, NY, 1984, incorporated herein by reference) established the structures shown in Figure 3. The megalomicins are 6-*O*-glycosides of erythromycin C with acetyl or propionyl groups esterified at the 3''' or 4''' hydroxyls of the mycarose sugar at the C-3-position. The C-6 sugar has been named "megosamine," although it had been identified 5 to 10 years earlier as L-rhodosamine or *N*-dimethyldaunosamine, deoxyamino sugars commonly present in the anthracycline antitumor drugs. The antibacterial potency, spectrum of activity, and toxicity (LD<sub>50</sub> acute, 7-7.5 g/kg s.c. or oral; subacute, >500 mg/kg) of the megalomicins is similar to that of erythromycin A.

The megalomicins have two modes of biological activity. As antibacterials, they act like the erythromycins, which inhibit protein synthesis at the translocation step by selective binding to the bacterial 50S ribosomal RNA. They also affect



protein trafficking in eukaryotic cells (see Bonay *et al.*, 1996, *J. Biol. Chem.* 271:3719-3726, incorporated herein by reference). Although the mechanism of action is not entirely clear, it appears to involve inhibition of vesicular transport between the medial and trans Golgi, resulting in under-sialylation of proteins. The megalomicins also strongly inhibit the ATP-dependent acidification of lysosomes *in vivo* (see Bonay *et al.*, 1997, *J. Cell. Sci.* 110:1839-1849, incorporated herein by reference) and cause an anomalous glycosylation of viral proteins, which may be responsible for their antiviral activity against herpes ( $\text{Tox}_{50}$ , 70-100  $\mu\text{M}$ ; see Alarcon *et al.*, 1984, *Antivir. Res.* 4:231-243, and Alarcon *et al.*, 1988, *FEBS Lett.* 231:207-211, both of which are incorporated herein by reference).

Strikingly, the megalomicins are potent antiparasitic agents, showing an  $\text{IC}_{50}$  of 1  $\mu\text{g/ml}$  in blocking intracellular replication of *Plasmodium falciparum* infected erythrocytes (see Bonay *et al.*, 1998, *Antimicrob. Agents Chemother.* 42:2668-2673, incorporated herein by reference). The megalomicins are effective against *Trypanosoma cruzi* and *T. brucei* ( $\text{IC}_{50}$ , 0.2-2  $\mu\text{g/ml}$ ) plus *Leishmania donovani* and *L. major* promastigotes ( $\text{IC}_{50}$ , 3 and 8  $\mu\text{g/ml}$ , respectively). Megalomicin is also active against the intracellular replicative, amastigote form of *T. cruzi*, completely preventing its replication in infected murine LLC/MK2 macrophages at a dose of 5  $\mu\text{g/ml}$ . Importantly, the effective drug concentration is 500-fold less than the acute  $\text{LD}_{50}$  in mammals, and there is no toxicity to BALB/c mice at doses (50 mg/kg) that are completely curative for *T. brucei* infections. Because the erythromycins do not have such activity, although azithromycin (Figure 3) has been reported to be an effective acute and prophylactic treatment for malaria caused by *P. vivax* and *P. falciparum* (see Taylor *et al.*, 1999, *Clin. Infect. Dis.* 28:74-81, incorporated herein by reference), the antiparasitic action of the megalomicins is unique and probably related to the presence of the deoxyamino sugar megosamine at C-6 (Figure 3). Consequently, the megalomicins could be developed into potent antimalarial drugs with a high therapeutic index and be active against *P. falciparum* and other species that are resistant to currently used classes of antimalarials. They also could lead to potent antiparasitic agents against leishmaniasis, trypanosomiasis, and Chagas' disease. In view of the widespread use of the erythromycins and their good oral availability plus the low mammalian toxicity of macrolides in general, the megalomicins could be used prophylactically

to combat malaria, and as fermentation products, the megalomicins should be relatively inexpensive to produce.

The megalomicins belong to the polyketide class of natural products whose members have diverse structural and pharmacological properties (see Monaghan  
5 and Tkacz, 1990, *Annu. Rev. Microbiol.* 44: 271, incorporated herein by reference). The megalomicins are assembled by polyketide synthases through successive condensations of activated coenzyme-A thioester monomers derived from small organic acids such as acetate, propionate, and butyrate. Active sites required for condensation include an acyltransferase (AT), acyl carrier protein  
10 (ACP), and beta-ketoacylsynthase (KS). Each condensation cycle results in a  $\beta$ -keto group that undergoes all, some, or none of a series of processing activities. Active sites that perform these reactions include a ketoreductase (KR), dehydratase (DH), and enoylreductase (ER). Thus, the absence of any beta-keto processing domain results in the presence of a ketone, a KR alone gives rise to a  
15 hydroxyl, a KR and DH result in an alkene, while a KR, DH, and ER combination leads to complete reduction to an alkane. After assembly of the polyketide chain, the molecule typically undergoes cyclization(s) and post-PKS modification (e.g. glycosylation, oxidation, acylation) to achieve the final active compound.

Macrolides such as erythromycin and megalomicin are synthesized by  
20 modular PKSs (see Cane *et al.*, 1998, *Science* 282: 63, incorporated herein by reference). For illustrative purposes, the PKS that produces the erythromycin polyketide (6-deoxyerythronolide B synthase or DEBS; see U.S. Patent No. 5,824,513, incorporated herein by reference) is shown in Figure 4. DEBS is the most characterized and extensively used modular PKS system. DEBS is  
25 particularly relevant to the present invention in that it synthesizes the same polyketide, 6-deoxyerythronolide B (6-dEB), synthesized by the megalomicin PKS. In modular PKS enzymes such as DEBS and the megalomicin PKS, the enzymatic steps for each round of condensation and reduction are encoded within a single "module" of the polypeptide (i.e., one distinct module for every  
30 condensation cycle). DEBS consists of a loading module and 6 extender modules and a chain terminating thioesterase (TE) domain within three extremely large polypeptides encoded by three open reading frames (ORFs, designated *eryAI*, *eryAII*, and *eryAIII*).



Each of the three polypeptide subunits of DEBS (DEBSI, DEBSII, and DEBSIII) contains 2 extender modules, DEBSI additionally contains the loading module. Collectively, these proteins catalyze the condensation and appropriate reduction of 1 propionyl CoA starter unit and 6 methylmalonyl CoA extender units. Modules 1, 2, 5, and 6 contain KR domains; module 4 contains a complete set, KR/DH/ER, of reductive and dehydratase domains; and module 3 contains no functional reductive domain. Following the condensation and appropriate dehydration and reduction reactions, the enzyme bound intermediate is lactonized by the TE at the end of extender module 6 to form 6-dEB.

More particularly, the loading module of DEBS consists of two domains, an acyl-transferase (AT) domain and an acyl carrier protein (ACP) domain. In other PKS enzymes, the loading module is not composed of an AT and an ACP but instead utilizes an inactivated KS, an AT, and an ACP. This inactivated KS is in most instances called KS<sup>Q</sup>, where the superscript letter is the abbreviation for the amino acid, glutamine, that is present instead of the active site cysteine required for activity. The AT domain of the loading module recognizes a particular acyl-CoA (propionyl for DEBS, which can also accept acetyl) and transfers it as a thiol ester to the ACP of the loading module. Concurrently, the AT on each of the extender modules recognizes a particular extender-CoA (methylmalonyl for DEBS) and transfers it to the ACP of that module to form a thioester. Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module migrates to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module 1 possesses an acyl-KS and a methylmalonyl ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading unit (elongation or extension). The growing polyketide chain is transferred from the ACP to the KS of the next module, and the process continues.

The polyketide chain, growing by two carbons each module, is sequentially passed as a covalently bound thiol ester from module to module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the

name polyketide arises. Commonly, however, the beta keto group of each two-carbon unit is modified just after it has been added to the growing polyketide chain but before it is transferred to the next module by either a KR, a KR plus a DH, or a KR, a DH, and an ER. As noted above, modules may contain additional enzymatic activities as well.

Once a polyketide chain traverses the final extender module of a PKS, it encounters the releasing domain or thioesterase found at the carboxyl end of most PKSs. Here, the polyketide is cleaved from the enzyme and cyclized. The resulting polyketide can be modified further by tailoring or modification enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, the final steps in conversion of 6-dEB to erythromycin A include the actions of a number of modification enzymes, such as: C-6 hydroxylation, attachment of mycarose and desosamine sugars, C-12 hydroxylation (which produces erythromycin C), and conversion of mycarose to cladinose via O-methylation, as shown in Figure 5.

With this overview of PKS and post-PKS modification enzymes, one can better appreciate the recombinant megalomicin biosynthetic genes provided by the invention and their function, as described in the following Section.

### Section III: The Megalomicin Biosynthetic Genes and Nucleic Acid Fragments

The megalomicin PKS was isolated and cloned by the following procedure. Genomic DNA was isolated from a megalomicin producing strain of *Micromonospora megalomicea* subsp. *nigra* (ATCC 27598), partially digested with a restriction enzyme, and cloned into a commercially available cosmid vector to produce a genomic library. This library was then probed with probe generated from the erythromycin biosynthetic genes as well as from cosmids identified as containing sequences homologous to erythromycin biosynthetic genes. This probing identified a set of cosmids, which were analyzed by DNA sequence analysis and restriction enzyme digestion, which revealed that the desired DNA had been isolated and that the entire PKS gene cluster was contained in overlapping segments on four of the cosmids identified. Figure 1 shows the cosmids, and the portions of the megalomicin biosynthetic gene cluster in the

insert DNA of the cosmids. Figure 1 shows that the complete megalomicin biosynthetic gene cluster is contained within the insert DNA of cosmids pKOS079-138B, pKOS079-124B, pKOS079-93D, and pKOS079-93A. Each of these cosmids has been deposited with the American Type Culture Collection in accordance with the terms of the Budapest Treaty (cosmid pKOS079-138B is available under accession no. ATCC \_\_\_\_; cosmid pKOS079-124B is available under accession no. ATCC \_\_\_\_; cosmid pKOS079-93D is available under accession no. ATCC \_\_\_\_; and cosmid pKOS079-93A is available under accession no. ATCC \_\_\_\_). Various additional reagents of the invention can be isolated from these cosmids. DNA sequence analysis was also performed on the various subclones of the invention, as described herein. Further analysis of these cosmids and subclones prepared from the cosmids facilitated the identification of the location of various megalomicin biosynthetic genes, including the ORFs encoding the PKS, modules encoded by those ORFs, and coding sequences for megalomicin modification enzymes. The location of these genes and modules is shown on Figure 2.

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the megalomicin PKS and other biosynthetic enzymes and other biosynthetic enzymes of *Micromonospora megalomicea* is shown herein merely to illustrate a preferred embodiment of the invention, and the invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences encoded by the DNA sequences shown herein merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following description of the various regions of the megalomicin PKS and the megalomicin modification

enzymes and corresponding coding sequences is provided. To facilitate description of the invention, reference to a PKS, protein, module, or domain herein can also refer to DNA compounds comprising coding sequences therefor and *vice versa*.

Also, unless otherwise indicated, reference to a heterologous PKS refers to a PKS  
5 or DNA compounds comprising coding sequences therefor from an organism other than *Micromonospora megalomicea*. In addition, reference to a PKS or its coding sequence includes reference to any portion thereof.

Thus, the invention provides DNA molecules in isolated (i.e., not pure, but existing in a preparation in an abundance and/or concentration not found in nature)  
10 and purified (i.e., substantially free of contaminating materials or substantially free of materials with which the corresponding DNA would be found in nature) form. The DNA molecules of the invention comprise one or more sequences that encode one or more domains (or fragments of such domains) of one or more modules in one or more of the ORFs of the megalomicin PKS and sequences that encode  
15 megalomicin modification enzymes from the megalomicin biosynthetic gene cluster. Examples of PKS domains include the KS, AT, DH, KR, ER, ACP, and TE domains of at least one of the 6 extender modules and loading module of the three proteins encoded by the three ORFs of the megalomicin PKS gene cluster. Examples of megalomicin modification enzymes include those that synthesize the  
20 mycarose, desosamine, and megosamine moieties, those that transfer those sugar moieties to the polyketide 6-dEB, those that hydroxylate the polyketide at C-6 and C-12, and those that acylate the sugar moieties.

In an especially preferred embodiment, the DNA molecule is a recombinant DNA expression vector or plasmid, as described in more detail in the  
25 following Section. Generally, such vectors can either replicate in the cytoplasm of the host cell or integrate into the chromosomal DNA of the host cell. In either case, the vector can be a stable vector (i.e., the vector remains present over many cell divisions, even if only with selective pressure) or a transient vector (i.e., the vector is gradually lost by host cells with increasing numbers of cell divisions).

30 The megalomicin PKS gene cluster comprises three ORFs (*megAI*, *megAII*, and *megAIII*). Each ORF encodes two extender modules of the PKS; the first ORF also encodes the loading module. Each extender module is composed of at least a KS, an AT, and an ACP domain. The locations of the various encoding regions of

these ORFs are shown in Figure 2 and described with reference to the sequence information below. The megalomicin PKS produces the polyketide known as 6-dEB, shown in Figure 4. In megalomicin-producing organisms, 6-dEB is converted to erythromycin C by a set of modification enzymes. Thus, 6-dEB is converted to erythronolide B by the *megF* gene product (a homolog of the *eryF* gene product), then to 3-alpha-mycarosyl-erythronolide B by the *megBV* gene product (a homolog of the *eryBV* gene product), then to erythromycin D by the *megCIII* gene product (a homolog of the *eryCIII* gene product), then to erythromycin C by the *megK* gene product (a homolog of the *eryK* gene product).

10 In addition to these modification enzymes, such megalomicin-producing organisms also contain the modification enzymes necessary for the biosynthesis of the desosamine and mycarose moieties that are similarly utilized in erythromycin biosynthesis, as shown in Figure 5. Megalomicin A contains the complete erythromycin C structure, and its biosynthesis additionally involves the formation of L-megosamine (L-rhodamine) and its attachment to the C-6 hydroxyl (Figures 3 and 5, inset), followed by acylation of the C-3''' and(or) C-4''' hydroxyls as the terminal steps. L-megosamine is the same as *N*-dimethyl-L-daunosamine; the daunosamine genes have been characterized from *Streptomyces peucetius* (see Colombo and Hutchinson, *J. Indust. Microbiol. Biotechnol.*, in press; Otten *et al.*, 1996, *J Bacteriol* 178:7316-7321, and references cited therein). Some of the rhodosamine genes also have been cloned and partially characterized from another anthracycline producing *Streptomyces* sp. (see Torkkell *et al.*, 1997, *Mol. Gen. Genet.* 256(2):203-209). Because the timing of the glycosylation with TDP-megosamine in relation to the addition of mycarose and desosamine to erythronolide B, plus the C-12 hydroxylation, is unknown, the pathway could involve a different order of glycosylation and C-12 hydroxylation steps than the one shown in Figure 5. Regardless, the megalomicin biosynthetic gene cluster contains the genes to make L-rhodamine and attach it to the correct macrolide substrate.

30 The biosynthetic pathways to make the glycosides desosamine, mycarose, and megosamine are shown in Figure 6. The present invention provides the genes for each biosynthetic pathway shown in this Figure, and these recombinant genetic

pathways can be used alone or in any combination to confer the pathway to a heterologous host.

The megalomicin PKS locus is similar to the *eryA* locus in size and organization. Most of the deoxysugar biosynthesis genes are homologs of the *eryB* mycarose and *eryC* desosamine biosynthesis and glycosyl attachment genes from *Saccharopolyspora erythraea* (see Summers *et al.*, 1997, *Microbiol. 143*:3251-3262; Haydock *et al.*, 1991, *Mol. Gen. Genet.* 230:120-128; Gaisser *et al.*, 1997, *Mol Gen Genet*, 256:239-251; Gaisser *et al.*, 1998, *Mol Gen Genet.* 257:78-88, incorporated herein by reference) or the *picC* homologs from the picromycin and narbomycin producer (see PCT patent publication No. 99/61599 and Xue *et al.*, 1998, *Proc. Nat. Acad. Sci. USA* 95, 12111-12116, incorporated herein by reference). The TDP-megosamine biosynthesis genes are homologs of the *dnm* genes (see Figure 5) and the pikromycin N-dimethyltransferase gene or its homologs reported in a cluster of L-rhodamine biosynthesis genes. The putative TDP-megosamine glycosyltransferase gene product (*geneX* in Figure 5) closely resembles the deduced products of the *eryBV*, *eryCIII*, *dnmS*, and pikromycin *desVII* genes, even though it recognizes different substrates than the products of each of these genes.

The following Table 1 shows the location of the genes in the *Micromonospora megalomicea* megalomicin biosynthetic pathway in the DNA sequence set forth in SEQ ID NO:1 (see also Figure 7; note some gene designations maybe different in Figure 7).

Table 1. Megalomicin Biosynthetic Gene Cluster  
*Micromonospora megalomicea* subsp. *nigra* (ATCC27598)

<u>Location</u>	<u>Description</u>
1..2451	sequence from cosmid pKOS079-138B
complement(1..144)	<i>megBVI</i> (or <i>megT</i> ), TDP-4-keto-6-deoxyglucose-
2,3-dehydratase	
928..2061	<i>megDVI</i> , TDP-4-keto-6-deoxyglucose 3,4-isomerase
2072..3382	<i>megDI</i> , TDP-megosaminyl transferase ( <i>eryCIII</i>
homolog)	
2452..40397	sequence of cosmid pKOS079-93D
3462..4634	<i>megG</i> (or <i>megY</i> ), mycarosyl acyltransferase
4651..5775	<i>megDII</i> , deoxysugar transaminase ( <i>eryCI</i> , <i>DnrJ</i>
	homolog)



	5822..6595	<i>megDIII</i> , TDP-daunosaminyl-N,N-
	dimethyltransferase	( <i>eryCVI</i> homolog)
5	6592..7197	<i>megDIV</i> , TDP-4-keto-6-deoxyglucose 3,5-epimerase
		( <i>eryBVII</i> , <i>dnmU</i> homolog)
	7220..8206	<i>megDV</i> , TDP-hexose 4-ketoreductase ( <i>eryBIV</i> ,
	<i>dnmV</i>	homolog)
10	complement(8228..9220)	<i>megBII-1</i> or <i>megDVII</i> , TDP-4-keto-L-6-deoxy-
	hexose 2,3-reductase	
	complement(9226..10479)	<i>megBV</i> , TDP-mycarosyl transferase
	complement(10483..11424)	<i>megBIV</i> , TDP-hexose 4-ketoreductase
	12181..22821	<i>megAI</i>
	12181..13791	Loading Module (L)
15	12505..13470	AT-L
	13576..13791	ACP-L
	13849..18207	Extender Module 1 (1)
	13849..15126	KS1
	15427..16476	AT1
20	17155..17694	KR1
	17947..18207	ACP1
	18268..22575	Extender Module 2 (2)
	18268..19548	KS2
	19876..20910	AT2
25	21517..22053	KR2
	22318..22575	ACP2
	22867..33555	<i>megAII</i>
	22957..27258	Extender Module 3 (3)
	22957..24237	KS3
30	24544..25581	AT3
	26230..26733	KR3 (inactive)
	26998..27258	ACP3
	27313..33312	Extender Module 4 (4)
	27393..28590	KS4
35	28897..29931	AT4
	29953..30477	DH4
	31396..32244	ER4
	32257..32799	KR4
	33052..33312	ACP4
40	33666..43271	<i>megAIII</i>
	33780..38120	Extender Module 5 (5)
	33780..35027	KS5
	35385..36419	AT5
	37068..37604	KR5
45	37860..38120	ACP5
	38187..42425	Extender Module 6 (6)
	38187..39470	KS6
	39795..40811	AT6
	40398..46641	sequences from cosmid pKOS079-93A

	41406..41936	KR6
	42168..42425	ACP6
	42585..43271	TE
	43268..44344	<i>megCII</i> , TDP-4-keto-6-deoxyglucose 3,4-isomerase
5	44355..45623	<i>megCIII</i> , TDP-desosaminy transferase
	45620..46591	<i>megBII</i> , TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase
	complement(46660..47403)	<i>megH</i> , TEII
	complement(47411..47980)	<i>megF</i> , C-6 hydroxylase

10

In a specific embodiment, the invention provides an isolated nucleic acid fragment comprising a nucleotide sequence encoding a domain of the megalomicin polyketide synthase or a megalomicin modification enzyme. The isolated nucleic acid fragment can be a DNA or a RNA. Preferably, the isolated  
 15 nucleic acid fragment is a recombinant DNA compound. A nucleotide sequence that is complementary to the nucleotide sequence encoding a domain of megalomicin PKS or a megalomicin modification enzyme is also provided.

The isolated nucleic acid fragment can comprise a single, multiple or all the open reading frame(s) (ORF) of the megalomicin PKS or the megalomicin  
 20 modification enzyme. Exemplary ORFs of megalomicin PKS include the ORFs of the *megAI*, *megAII* and *megAIII* genes. The isolated nucleic acids of the invention also include nucleic acids that encode one or more domains and one or more modules of the megalomicin PKS. Exemplary domains of the megalomicin PKS include a TE domain, a KS domain, an AT domain, an ACP domain, a KR  
 25 domain, a DH domain and an ER domain. In a preferred embodiment, the nucleic acid comprises the coding sequence for a loading module, a thioesterase domain, and all six extender modules of the megalomicin PKS.

Megalomicin modification enzymes include those enzymes involved in the conversion of 6-DEB into a megalomicin such as the enzymes encoded by *megF*,  
 30 *megBV*, *megCIII*, *megK*, *megDI* and *megG* (or *megY*). Megalomicin modification enzymes also include those enzymes involved in the biosynthesis of mycarose, megosamine or desosamine, which are used as biosynthetic intermediates in the biosynthesis of various megalomicin species and other related polyketides. The enzymes that are involved in biosynthesis of mycarose, megosamine or  
 35 desosamine are described in Figures 5 and 10. The megalomicin PKS and megalomicin modification enzymes are collectively referred to as megalomicin



biosynthetic enzymes; the genes encoding such enzymes are collectively referred to as megalomicin biosynthetic genes; and nucleic acids that comprise a portion of or entire megalomicin biosynthetic genes are collectively referred to as megalomicin biosynthetic nucleic acid(s).

5 In specific embodiments, the megalomicin biosynthetic nucleic acids comprise the sequence of SEQ ID NO:1, or the coding regions thereof, or nucleotide sequences encoding, in whole or in part, a megalomicin biosynthetic enzyme protein. The isolated nucleic acids typically consists of at least 25 (continuous) nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200  
10 nucleotides of megalomicin biosynthetic nucleic acid sequence, or a full-length megalomicin biosynthetic coding sequence. In another embodiment, the nucleic acids are smaller than 35, 200, or 500 nucleotides in length. Nucleic acids can be single or double stranded. Nucleic acids that hybridize to or are complementary to the foregoing sequences, in particular the inverse complement to nucleic acids that  
15 hybridize to the foregoing sequences (*i.e.*, the inverse complement of a nucleic acid strand has the complementary sequence running in reverse orientation to the strand so that the inverse complement would hybridize without mismatches to the nucleic acid strand) are also provided. In specific aspects, nucleic acids are provided which comprise a sequence complementary to (specifically are the  
20 inverse complement of) at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a megalomicin biosynthetic gene.

The megalomicin biosynthetic nucleic acids provided herein include those with nucleotide sequences encoding substantially the same amino acid sequences as found in native megalomicin biosynthetic enzyme proteins, and those encoding  
25 amino acid sequences with functionally equivalent amino acids, as well as megalomicin biosynthetic enzyme derivatives or analogs as described in Section IV.

Some regions within the megalomicin PKS genes are highly homologous or identical to one another, as can be readily identified by an analysis of the  
30 sequence. The coding sequence for the KS and AT domains of module 2 shares significant identity with the coding sequence for the KS and AT domains of module 6. This sequence homology or identity at the nucleic acid, *e.g.*, DNA, level can render the nucleic acid unstable in certain host cells. To improve the stability

of the nucleic acids comprising a portion or the entire megalomicin PKS genes and megalomicin modification enzyme genes, the nucleic acid or DNA sequences can be changed to reduce or abolish the sequence homology or identity. Preferably, the DNA codons of homologous regions within the PKS or the megalomicin  
5 modification enzyme coding sequence are changed to reduce or abolish the sequence homology or identity without changing the amino acid sequences encoded by said changed DNA codons (see the examples below). The stability of the nucleic acid or DNA can also be improved by codon changes that reduce or abolish the sequence homology or identity while also changing the amino acid  
10 sequence, provided that the amino acid sequence change(s) does not substantially change the desired activity of the encoded megalomicin PKS. Thus, for example, one can simply substitute for the *megAIII* ORF an ORF from *eryAIII*, *oleAIII*, *picAIII*, or *picAIV* genes.

The recombinant DNA compounds of the invention that encode the  
15 megalomicin PKS and modification proteins or portions thereof are useful in a variety of applications. While many of these applications relate to the heterologous expression of the megalomicin biosynthetic genes or the construction of hybrid PKS enzymes, many useful applications involve the natural megalomicin producer *Micromonospora megalomicea*. For example, one can use the recombinant DNA  
20 compounds of the invention to disrupt the megalomicin biosynthetic genes by homologous recombination in *Micromonospora megalomicea*. The resulting host cell is a preferred host cell for making polyketides modified by oxidation, hydroxylation, glycosylation, and acylation in a manner similar to megalomicin, because the genes that encode the proteins that perform these reactions are of  
25 course present in the host cell, and because the host cell does not produce megalomicin that could interfere with production or purification of the polyketide of interest.

One illustrative recombinant host cell provided by the present invention expresses a recombinant megalomicin PKS in which the module 1 KS domain is  
30 inactivated by deletion or other mutation. In a preferred embodiment, the inactivation is mediated by a change in the KS domain that renders it incapable of binding substrate (called a KS1° mutation). In a particularly preferred embodiment, this inactivation is rendered by a mutation in the codon for the active

site cysteine that changes the codon to another codon, such as an alanine codon. Such constructs are especially useful when placed in translational reading frame with extender modules 1 and 2 of a megalomicin or the corresponding modules of another PKS. The utility of these constructs is that host cells expressing, or cell  
5 free extracts containing, a PKS comprising the protein encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare a polyketide of interest. See U.S. patent application Serial No. 09/492,773, filed 27 Jan. 2000, and PCT patent publication No. 00/44717, both of which are incorporated herein by reference. Such KS1° constructs of the invention are useful  
10 in the production of 13-substituted-megalomicin compounds in *Micromonospora megalomicea* host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl.

In a variant of this embodiment, one can employ a megalomicin PKS in  
15 which the ACP domain of module 1 has been rendered inactive. In another embodiment, one can delete the loading domain of the megalomicin PKS and provide monoketide substrates for processing by the remainder of the PKS.

The compounds of the invention can also be used to construct recombinant host cells of the invention in which coding sequences for one or more domains or  
20 modules of the megalomicin PKS or for another megalomicin biosynthetic gene have been deleted by homologous recombination with the *Micromonospora megalomicea* chromosomal DNA. Those of skill in the art will appreciate that the compounds used in the recombination process are characterized by their homology with the chromosomal DNA and not by encoding a functional protein due to their  
25 intended function of deleting or otherwise altering portions of chromosomal DNA. For this and a variety of other applications, the compounds of the present invention include not only those DNA compounds that encode functional proteins but also those DNA compounds that are complementary or identical to any portion of the megalomicin biosynthetic genes.

30 Thus, the invention provides a variety of modified *Micromonospora megalomicea* host cells in which one or more of the megalomicin biosynthetic genes have been mutated or disrupted. Transformation systems for *M. megalomicea* have been described by Hasegawa *et al.*, 1991, *J. Bacteriol.*

173:7004-11; and Takada *et al.*, 1994, *J. Antibiot.* 47:1167-1170, both of which are incorporated herein by reference. These cells are especially useful when it is desired to replace the disrupted function with a gene product expressed by a recombinant DNA expression vector. While such expression vectors of the invention are described in more detail in the following Section, those of skill in the art will appreciate that the vectors have application to *M. megalomicea* as well. Such *M. megalomicea* host cells can be preferred host cells for expressing megalomicin derivatives of the invention. Particularly preferred host cells of this type include those in which the coding sequence for the loading module has been mutated or disrupted, those in which one or more of any of the PKS gene ORFs has been mutated or disrupted, and/or those in which the genes for one or more modification (glycosylation, acylation, hydroxylation) have been mutated or disrupted.

While the present invention provides many useful compounds having application to, and recombinant host cells derived from, *Micromonospora megalomicea*, many important applications of the present invention relate to the heterologous expression of all or a portion of the megalomicin biosynthetic genes in cells other than *M. megalomicea*, as described in Section V.

#### 20 Section IV: The Megalomicin Biosynthetic Enzymes and Antibodies Recognizing such Enzymes

In another specific embodiment, the invention provides a substantially purified polypeptide, which is encoded by a nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The polypeptide can comprise a single domain, multiple domains or a full-length megalomicin PKS or megalomicin modification enzyme. Functional fragments, analogs or derivatives of the megalomicin PKS or megalomicin modification enzyme polypeptides are also provided. Preferably, such fragments, analogs or derivatives can be recognized an antibody raised against a megalomicin PKS or megalomicin modification enzyme. Also preferably, such fragments, analogs or derivatives comprise an amino acid sequence that has at least 60% identity, more preferably at least 90% identity to their wild type counterparts.

An exemplary nucleotide sequence encoding, and the corresponding amino acid sequence of, a megalomicin biosynthetic enzyme is disclosed in SEQ ID NO:1. Homologs (*e.g.*, nucleic acids of the above-listed genes of species other than *Micromonospora megalomicea*) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular sequence provided as a probe using methods well known in the art for nucleic acid hybridization and cloning (*e.g.*, as described in Section III) in accordance with the methods of the present invention.

The megalomicin biosynthetic enzyme proteins, or domains thereof, of the present invention can be obtained by methods well known in the art for protein purification and recombinant protein expression in accordance with the methods of the present invention. For recombinant expression of one or more of the proteins, the nucleic acid containing all or a portion of the nucleotide sequence encoding the protein can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein coding sequence. Transcriptional and translational signals can be supplied by the native promoter for a megalomicin biosynthetic gene and/or flanking regions.

A variety of host-vector systems may be utilized to express the protein coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, and the like); insect cell systems infected with virus (*e.g.* baculovirus); microorganisms such as yeast containing yeast vectors; or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their properties. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used.

In a specific embodiment, a vector is used that comprises a promoter operably linked to nucleic acid sequences encoding a megalomicin biosynthetic enzyme, or a domain, fragment, derivative or homolog, thereof, one or more origins of replication, and optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

Expression vectors containing the sequences of interest can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or

absence of "marker" gene function, and (c) expression of the inserted sequences. In the first approach, megalomicin biosynthetic nucleic acid sequences can be detected by nucleic acid hybridization to probes comprising sequences homologous and complementary to the inserted sequences. In the second  
5 approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" functions (*e.g.*, binding to an anti-megalomicin biosynthetic enzyme antibody, resistance to antibiotics, occlusion body formation in baculovirus, and the like) caused by insertion of the sequences of interest in the vector. For example, if a megalomicin biosynthetic  
10 gene, or portion thereof, is inserted within the marker gene sequence of the vector, recombinants containing the megalomicin biosynthetic gene fragment will be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying for the megalomicin biosynthetic gene products expressed by the recombinant vector. Such assays can  
15 be based, for example, on the physical or functional properties of the interacting species in *in vitro* assay systems, *e.g.*, megalomicin synthesis activity, immunoreactivity to antibodies specific for the protein.

Once recombinant megalomicin biosynthetic genes or nucleic acids are identified, several methods known in the art can be used to propagate them in  
20 accordance with the methods of the present invention. Once a suitable host system and growth conditions have been established, recombinant expression vectors can be propagated and amplified in quantity. As previously described, the expression vectors or derivatives which can be used include, but are not limited to: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such  
25 as baculovirus, yeast vectors; bacteriophage vectors such as lambda phage; and plasmid and cosmid vectors.

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies or processes the expressed proteins in the specific fashion desired. Expression from certain promoters can be elevated in the  
30 presence of certain inducers; thus expression of the genetically-engineered megalomicin biosynthetic enzymes may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.* glycosylation, phosphorylation, and



the like) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein is achieved. For example, expression in a bacterial system can be used to produce an unglycosylated core protein, while expression in mammalian cells ensures  
5 "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions to different extent.

In particular, megalomicin biosynthetic enzyme derivatives can be made by altering their sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding  
10 sequences, other DNA sequences which encode substantially the same amino acid sequence as an megalomicin biosynthetic gene can be used in the practice of the present invention. These include but are not limited to nucleotide sequences comprising all or portions of megalomicin biosynthetic genes that are altered by the substitution of different codons that encode the amino acid residue within the  
15 sequence, thus producing a silent change. Likewise, the megalomicin biosynthetic enzyme derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of megalomicin biosynthetic enzymes, including altered sequences in which functionally equivalent amino acid residues are substituted for residues  
20 within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example,  
25 the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and  
30 glutamic acid.

In a specific embodiment of the invention, the nucleic acids encoding proteins and proteins consisting of or comprising a domain or a fragment of megalomicin biosynthetic enzyme consisting of at least 6 (continuous) amino

acids are provided. In other embodiments, the domain or fragment consists of at least 10, 20, 30, 40, or 50 amino acids of a megalomicin biosynthetic enzyme. In specific embodiments, such domains or fragments are not larger than 35, 100 or 200 amino acids. Derivatives or analogs of megalomicin biosynthetic enzyme  
5 include but are not limited to molecules comprising regions that are substantially homologous to megalomicin biosynthetic enzyme in various embodiments, at least 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% identity over an amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art in  
10 accordance with the methods of the present invention or whose encoding nucleic acid is capable of hybridizing to a sequence encoding a megalomicin biosynthetic enzyme under stringent, moderately stringent, or nonstringent conditions.

The megalomicin biosynthetic enzyme domains, derivatives and analogs of the invention can be produced by various methods known in the art in accordance  
15 with the methods of the present invention. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned megalomicin biosynthetic gene sequence can be modified by any of numerous strategies known in the art (Sambrook et al., 1990, *Molecular Cloning, A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor,  
20 New York) in accordance with the methods of the present invention. The sequences can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*.

Additionally, the megalomicin biosynthetic enzyme-encoding nucleotide  
25 sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy pre-existing ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used in accordance with the methods of the present invention,  
30 including but not limited to, chemical mutagenesis and *in vitro* site-directed mutagenesis (Hutchinson et al., *J. Biol. Chem.* 253:6551-6558 (1978)), use of TAB® linkers (Pharmacia), and the like.



Once a recombinant cell expressing a megalomicin biosynthetic enzyme protein, or a domain, fragment or derivative thereof, is identified, the individual gene product can be isolated and analyzed. This is achieved by assays based on the physical and/or functional properties of the protein, including, but not limited to, radioactive labeling of the product followed by analysis by gel electrophoresis, immunoassay, cross-linking to marker-labeled product, and the like.

The megalomicin biosynthetic enzyme proteins may be isolated and purified by standard methods known in the art or recombinant host cells expressing the complexes or proteins in accordance with the methods of the invention, including but not restricted to column chromatography (e.g., ion exchange, affinity, gel exclusion, reversed-phase high pressure, fast protein liquid, and the like), differential centrifugation, differential solubility, or by any other standard technique used for the purification of proteins. Functional properties may be evaluated using any suitable assay known in the art in accordance with the methods of the present invention.

Alternatively, once a megalomicin biosynthetic enzyme or its domain or derivative is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the gene which encodes it. As a result, the protein or its domain or derivative can be synthesized by standard chemical methods known in the art in accordance with the methods of the present invention (see Hunkapiller et al, *Nature* 310:105-111 (1984)).

Manipulations of megalomicin biosynthetic enzymes may be made at the protein level. Included within the scope of the invention are megalomicin biosynthetic enzyme domains, derivatives or analogs or fragments, which are differentially modified during or after translation, e.g., by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, and the like. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH<sub>4</sub>, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, and the like.

In specific embodiments, the megalomicin biosynthetic enzymes are modified to include a fluorescent label. In other specific embodiments, the megalomicin biosynthetic enzyme is modified to have a heterofunctional reagent, such heterofunctional reagents can be used to crosslink the members of the  
5 complex.

In addition, domains, analogs and derivatives of a megalomicin biosynthetic enzyme can be chemically synthesized. For example, a peptide corresponding to a portion of a megalomicin biosynthetic enzyme, which comprises the desired domain or which mediates the desired activity *in vitro* can  
10 be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the megalomicin biosynthetic enzyme sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, alpha-amino isobutyric acid, 4-aminobutyric acid,  
15 2-aminobutyric acid, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine,  $\beta$ -alanine, fluoro-amino acids, designer amino acids such as  $\beta$ -methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino  
20 acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In cases where natural products are suspected of being mutant or are isolated from new species, the amino acid sequence of the megalomicin biosynthetic enzyme isolated from the natural source, as well as those expressed *in*  
25 *vitro*, or from synthesized expression vectors *in vivo* or *in vitro*, can be determined from analysis of the DNA sequence, or alternatively, by direct sequencing of the isolated protein. Such analysis may be performed by manual sequencing or through use of an automated amino acid sequenator.

The megalomicin biosynthetic enzyme proteins may also be analyzed by  
30 hydrophilicity analysis (Hopp and Woods, *Proc. Natl. Acad. Sci. USA* 78:3824-3828 (1981)). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the proteins, and help predict their orientation in designing substrates for experimental manipulation, such as in binding

experiments, antibody synthesis, and the like. Secondary structural analysis can also be done to identify regions of the megalomicin biosynthetic enzyme that assume specific structures (Chou and Fasman, *Biochemistry* 13:222-23 (1974)). Manipulation, translation, secondary structure prediction, hydrophilicity and hydrophobicity profiles, open reading frame prediction and plotting, and determination of sequence homologies, can be accomplished using computer software programs available in the art.

Other methods of structural analysis including but not limited to X-ray crystallography (Engstrom, *Biochem. Exp. Biol.* 11:7-13 (1974)), mass spectroscopy and gas chromatography (Methods in Protein Science, J. Wiley and Sons, New York, 1997), and computer modeling (Fletterick and Zoller, eds., 1986, Computer Graphics and Molecular Modeling, In: *Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, New York) can also be employed.

The invention also provides an antibody, or a fragment or derivative thereof, which immuno-specifically binds to a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. In a specific embodiment, an antibody which immuno-specifically binds to a domain of the megalomicin biosynthetic enzyme encoded by a nucleic acid that hybridizes to a nucleic acid having the nucleotide sequence set forth in the SEQ. ID NO:1, or a fragment or derivative of said antibody containing the binding domain thereof is provided. Preferably, the antibody is a monoclonal antibody.

The megalomicin biosynthetic enzyme protein and domains, fragments, homologs and derivatives thereof may be used as immunogens to generate antibodies which immunospecifically bind such immunogens. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

Various procedures known in the art may be used for the production of polyclonal antibodies to a megalomicin biosynthetic enzyme protein of the invention, its domains, derivatives, fragments or analogs in accordance with the methods of the present invention.

For production of the antibody, various host animals can be immunized by injection with the native megalomicin biosynthetic enzyme protein or a synthetic

version, or a derivative of the foregoing, such as a cross-linked megalomicin biosynthetic enzyme. Such host animals include but are not limited to rabbits, mice, rats, and the like. Various adjuvants can be used to increase the immunological response, depending on the host species, and include but are not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, and potentially useful human adjuvants such as bacille Calmette-Guerin (BCG) and corynebacterium parvum.

For preparation of monoclonal antibodies directed towards a megalomicin biosynthetic enzyme or domains, derivatives, fragments or analogs thereof, any technique that provides for the production of antibody molecules by continuous cell lines in culture may be used. Such techniques include but are not restricted to the hybridoma technique originally developed by Kohler and Milstein (*Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983)), and the EBV hybridoma technique to produce human monoclonal antibodies (Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). In an additional embodiment, monoclonal antibodies can be produced in germ-free animals (WO89/12690). Human antibodies may be used and can be obtained by using human hybridomas (Cote et al., *Proc. Natl. Acad. Sci. USA* 80:2026-2030 (1983)) or by transforming human B cells with EBV virus *in vitro* (Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). Techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci. USA* 81:6851-6855 (1984); Neuberger et al., *Nature* 312:604-608 (1984); Takeda et al., *Nature* 314:452-454 (1985)) by splicing the genes from a mouse antibody molecule specific for the megalomicin biosynthetic enzyme protein together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention.

Techniques described for the production of single chain antibodies (U.S. patent 4,946,778) can be adapted to produce megalomicin biosynthetic enzyme-specific single chain antibodies. An additional embodiment utilizes the techniques described for the construction of Fab expression libraries (Huse et al., *Science*

246:1275-1281 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for megalomicin biosynthetic enzyme, or domains, derivatives, or analogs thereof. Non-human antibodies can be "humanized" by known methods (*see, e.g.*, U.S. Patent No. 5,225,539).

5           Antibody fragments that contain the idiotypes of a megalomicin biosynthetic enzyme can be generated by techniques known in the art in accordance with the methods of the present invention. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments that  
10       can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment, the Fab fragments that can be generated by treating the antibody molecular with papain and a reducing agent, and Fv fragments.

          In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art in accordance with the methods of  
15       the present invention, *e.g.*, ELISA (enzyme-linked immunosorbent assay). To select antibodies specific to a particular domain of the megalomicin biosynthetic enzyme, one may assay generated hybridomas for a product that binds to the fragment of a megalomicin biosynthetic enzyme that contains such a domain.

          The foregoing antibodies can be used in methods known in the art relating  
20       to the localization and/or quantitation of megalomicin biosynthetic enzyme proteins, *e.g.*, for imaging these proteins or measuring levels thereof in samples, in accordance with the methods of the present invention.

#### Section V: Heterologous Expression of the Megalomicin Biosynthetic Genes

25           In one important embodiment, the invention provides methods for the heterologous expression of one or more of the megalomicin biosynthetic genes and recombinant DNA expression vectors useful in the method. For purposes of the invention, any host cell other than *Micromonospora megalomicea* is a heterologous host cell. Thus, included within the scope of the invention in  
30       addition to isolated nucleic acids encoding domains, modules, or proteins of the megalomicin PKS and modification enzymes, are recombinant expression vectors that include such nucleic acids. The term expression vector refers to a nucleic acid that can be introduced into a host cell or cell-free transcription and translation

system. An expression vector can be maintained permanently or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a promoter that drives expression of an RNA, which typically is translated into a polypeptide in the cell or cell extract. For efficient translation of RNA into protein, the expression vector also typically contains a ribosome-binding site sequence positioned upstream of the start codon of the coding sequence of the gene to be expressed. Other elements, such as enhancers, secretion signal sequences, transcription termination sequences, and one or more marker genes by which host cells containing the vector can be identified and/or selected, may also be present in an expression vector. Selectable markers, i.e., genes that confer antibiotic resistance or sensitivity, are preferred and confer a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium.

The various components of an expression vector can vary widely, depending on the intended use of the vector and the host cell(s) in which the vector is intended to replicate or drive expression. Expression vector components suitable for the expression of genes and maintenance of vectors in *E. coli*, yeast, *Streptomyces*, and other commonly used cells are widely known and commercially available. For example, suitable promoters for inclusion in the expression vectors of the invention include those that function in eucaryotic or procaryotic host cells. Promoters can comprise regulatory sequences that allow for regulation of expression relative to the growth of the host cell or that cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus. For *E. coli* and certain other bacterial host cells, promoters derived from genes for biosynthetic enzymes, antibiotic-resistance conferring enzymes, and phage proteins can be used and include, for example, the galactose, lactose (*lac*), maltose, tryptophan (*trp*), beta-lactamase (*bla*), bacteriophage lambda PL, and T5 promoters. In addition, synthetic promoters, such as the *tac* promoter (U.S. Patent No. 4,551,433), can also be used.

Thus, recombinant expression vectors contain at least one expression system, which, in turn, is composed of at least a portion of the megalomicin PKS and/or other megalomicin biosynthetic gene coding sequences operably linked to a



promoter and optionally termination sequences that operate to effect expression of the coding sequence in compatible host cells. The host cells are modified by transformation with the recombinant DNA expression vectors of the invention to contain the expression system sequences either as extrachromosomal elements or  
5 integrated into the chromosome. The resulting host cells of the invention are useful in methods to produce PKS and post-PKS modification enzymes as well as polyketides and antibiotics and other useful compounds derived therefrom.

Preferred host cells for purposes of selecting vector components for expression vectors of the present invention include fungal host cells such as yeast  
10 and procaryotic host cells such as *E. coli* and *Streptomyces*, but mammalian host cells can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce polyketides, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality. Provision of such enzymes is  
15 described, for example, in PCT publication Nos. WO 97/13845 and 98/27203, each of which is incorporated herein by reference. Particularly preferred host cells for purposes of the present invention are *Streptomyces* and *Saccharopolyspora* host cells, as discussed in greater detail below.

In a preferred embodiment, the expression vectors of the invention are  
20 used to construct a heterologous recombinant *Streptomyces* host cell that expresses a recombinant PKS of the invention. *Streptomyces* is a convenient host for expressing polyketides, because polyketides are naturally produced in certain *Streptomyces* species, and *Streptomyces* cells generally produce the precursors needed to form the desired polyketide. Those of skill in the art will recognize that,  
25 if a *Streptomyces* host cell produces any portion of a PKS enzyme or produces a polyketide modification enzyme, the recombinant vector need drive expression of only those genes constituting the remainder of the desired PKS enzyme or other polyketide-modifying enzymes. Thus, such a vector may comprise only a single ORF, with the desired remainder of the polypeptides constituting the PKS  
30 provided by the genes on the host cell chromosomal DNA.

If a *Streptomyces* or other host cell ordinarily produces polyketides, it may be desirable to modify the host so as to prevent the production of endogenous polyketides prior to its use to express a recombinant PKS of the invention. Such

modified hosts include *S. coelicolor* CH999 and similarly modified *S. lividans* described in U.S. Patent No. 5,672,491, and PCT publication Nos. WO 95/08548 and WO 96/40968, incorporated herein by reference. In such hosts, it may not be necessary to provide enzymatic activities for all of the desired post-translational  
5 modifications of the enzymes that make up the recombinantly produced PKS, because the host naturally expresses such enzymes. In particular, these hosts generally contain holo-ACP synthases that provide the phosphopantotheinyl residue needed for functionality of the PKS.

The invention provides a wide variety of expression vectors for use in  
10 *Streptomyces*. The replicating expression vectors of the present invention include, for example and without limitation, those that comprise an origin of replication from a low copy number vector, such as SCP2\* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and  
15 Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and pSG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, *Plasmid* 30: 131-140, each of which is  
20 incorporated herein by reference). For non-replicating and integrating vectors and generally for any vector, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and pBR. For phage based vectors, the  
25 phage phiC31 and its derivative KC515 can be employed (see Hopwood *et al.*, *supra*). Also, plasmid pSET152, plasmid pSAM, plasmids pSE101 and pSE211, all of which integrate site-specifically in the chromosomal DNA of *S. lividans*, can be employed for purposes of the present invention.

The *Streptomyces* recombinant expression vectors of the invention  
30 typically comprise one or more selectable markers, including antibiotic resistance conferring genes selected from the group consisting of the *ermE* (confers resistance to erythromycin and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4*

(confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes. Alternatively, several polyketides are naturally colored, and this characteristic can provide a built-in marker for identifying cells.

Megalomicins are currently produced only by the relatively genetically intractable host *Micromonospora megalomicinea*. This bacteria has not been commonly used in the fermentation industry for the large-scale production of antibiotics, and methods for high level production of megalomicin and its analogs are needed. In contrast, the streptomycete bacteria have been widely used for almost 50 years and are excellent hosts for production of megalomicin and its analogs. *Streptomyces lividans* and *S. coelicolor* have been developed for the expression of heterologous PKS systems. These organisms can stably maintain cloned heterologous PKS genes, express them at high levels under controlled conditions, and modify the corresponding PKS proteins (e.g., phosphopantotheinylation) so that they are capable of production of the polyketide they encode. Furthermore, these hosts contain the necessary pathways to produce the substrates required for polyketide synthesis; e.g. propionyl-CoA and methylmalonyl-CoA. A wide variety of cloning and expression vectors are available for these hosts, as are methods for the introduction and stable maintenance of large segments of foreign DNA. Relative to *Micromonospora* spp., *S. lividans* and *S. coelicolor* grow well on a number of media and have been adapted for high level production of polyketides in fermentors. If production levels are low, a number of rational approaches are available to improve yield (see Hosted and Baltz, 1996, *Trends Biotechnol.* 14(7):245-50, incorporated herein by reference). Empirical methods to increase the titers of these macrolides, long since proven effective for numerous bacterial polyketides, can also be employed.

Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which have been modified so as not to produce the polyketide actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent Nos. 5,830,750 and 6,022,731 and U.S. patent application Serial No. 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. These vectors are

particularly preferred in that they contain promoters compatible with numerous and diverse *Streptomyces* spp. Particularly useful promoters for *Streptomyces* host cells include those from PKS gene clusters that result in the production of polyketides as secondary metabolites, including promoters from aromatic (Type II) PKS gene clusters. Examples of Type II PKS gene cluster promoters are *act* gene promoters and *tcm* gene promoters; an example of a Type I PKS gene cluster promoter are the promoters of the spiramycin PKS genes and DEBS genes. The present invention also provides the megalomicin biosynthetic gene promoters in recombinant form. These promoters can be used to drive expression of the megalomicin biosynthetic genes or any other coding sequence of interest in host cells in which the promoter functions, particularly *Micromonospora megalomicea* and generally any *Streptomyces* species.

As described above, particularly useful control sequences are those that alone or together with suitable regulatory systems activate expression during transition from growth to stationary phase in the vegetative mycelium. The promoter contained in the aforementioned plasmid pRM5, i.e., the *actII/actIII* promoter pair and the *actII-ORF4* activator gene, is particularly preferred. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actII-ORF4* gene described above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, supra).

To provide a preferred host cell and vector for purposes of the invention, the megalomicin biosynthetic genes are placed on a recombinant expression vector and transferred to the non-macrolide producing hosts *Streptomyces lividans* K4-114 and *S. coelicolor* CH999. Transformation of *S. lividans* K4-114 or *S. coelicolor* CH999 with this expression vector results in a strain which produces

detectable amounts of megalomicin as determined by analysis of extracts by LC/MS. As noted above, the present invention also provides recombinant DNA compounds in which the encoded megalomicin module 1 KS domain is inactivated (the KS1° mutation). The introduction into *Streptomyces lividans* or *S. coelicolor* of a recombinant expression vector of the invention that encodes a megalomicin PKS with a KS1° domain produces a host cell useful for making polyketides by a process known as diketide feeding. The resulting host cells can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare megalomicin derivatives. Such cells of the invention are especially useful in the production of 13-substituted-6-deoxyerythronolide B compounds in recombinant host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl. In a preferred embodiment, the meg PKS is produced from a recombinant construct in which the *megAIII* gene has been altered to abolish the regions of identical coding sequence it otherwise shares with the *megAI* gene, or a hybrid PKS is employed in which the *megAIII* gene product has been replaced by the *oleAIII* gene product. Recombinant *oleAIII* genes are described in, for example, PCT patent publication No. 00/026349 and U.S. patent application Serial No. 09/428,517, filed 28 Oct. 1999, both of which are incorporated herein by reference.

The recombinant host cells of the invention can express all of the megalomicin biosynthetic genes or only a subset of the same. For example, if only the genes for the megalomicin PKS are expressed in a host cell that otherwise does not produce polyketide modifying enzymes that can act on the polyketide produced, then the host cell produces unmodified polyketides, called macrolide aglycones. Such macrolide aglycones can be hydroxylated and glycosylated by adding them to the fermentation of a strain such as, for example, *Streptomyces antibioticus* or *Saccharopolyspora erythraea*, that contains the requisite modification enzymes.

There are a wide variety of diverse organisms that can modify macrolide aglycones to provide compounds with, or that can be readily modified to have, useful activities. For example, as shown in Figure 5, *Saccharopolyspora erythraea* can convert 6-dEB to a variety of useful compounds. The erythronolide 6-dEB is



converted by the *eryF* gene product to erythronolide B, which is, in turn, glycosylated by the *eryBV* gene product to obtain 3-O-mycarosylerythronolide B, which contains L-mycarose at C-3. The *eryCIII* gene product then converts this compound to erythromycin D by glycosylation with D-desosamine at C-5.

- 5 Erythromycin D, therefore, differs from 6-dEB through glycosylation and by the addition of a hydroxyl group at C-6. Erythromycin D can be converted to erythromycin B in a reaction catalyzed by the *eryG* gene product by methylating the L-mycarose residue at C-3. Erythromycin D is converted to erythromycin C by the addition of a hydroxyl group at C-12 in a reaction catalyzed by the *eryK* gene product.
- 10 Erythromycin A is obtained from erythromycin C by methylation of the mycarose residue in a reaction catalyzed by the *eryG* gene product. The unmodified megalomicin compounds provided by the present invention, such as, for example, the 6-dEB or 6-dEB analogs, produced in *Streptomyces lividans*, can be provided to cultures of *S. erythraea* and converted to the corresponding
- 15 derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in the examples below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol.* 164(1): 425-433). Also, one can employ other mutant strains, such as
- 20 *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production.

Moreover, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described

25 above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be chemically altered after fermentation. Thus, *Streptomyces venezuelae*, which produces picromycin, contains enzymes that can transfer a desosaminyl group to the C-5 hydroxyl and a

30 hydroxyl group to the C-12 position. In addition, *S. venezuelae* contains a glucosylation activity that glucosylates the 2'-hydroxyl group of the desosamine sugar. This latter modification reduces antibiotic activity, but the glucosyl residue is removed by enzymatic action prior to release of the polyketide from the cell.



Another organism, *S. narbonensis*, contains the same modification enzymes as *S. venezuelae*, except the C-12 hydroxylase. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. narbonensis* and *S. venezuelae*.

Other organisms suitable for making compounds of the invention include *Micromonospora megalomicea* (discussed above), *Streptomyces antibioticus*, *S. fradiae*, and *S. thermotolerans*. *S. antibioticus* produces oleandomycin and contains enzymes that hydroxylate the C-6 and C-12 positions, glycosylate the C-3 hydroxyl with oleandrose and the C-5 hydroxyl with desosamine, and form an epoxide at C-8-C-8a. *S. fradiae* contains enzymes that glycosylate the C-5 hydroxyl with mycaminose and then the 4'-hydroxyl of mycaminose with mycarose, forming a disaccharide. *S. thermotolerans* contains the same activities as *S. fradiae*, as well as acylation activities. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. antibioticus*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *megAI*, *megAII*, and *megAIII* genes with one or more deletions and/or insertions, including replacements of a *megA* gene fragment with a gene fragment from a heterologous PKS gene (as discussed in the next Section), can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Streptomyces antibioticus*, *S. venezuelae*, *S. narbonensis*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans*.

A number of erythromycin high-producing strains of *Saccharopolyspora erythraea* and *Streptomyces fradiae* have been developed, and in a preferred embodiment, the megalomicin PKS and/or other megalomicin biosynthetic genes are introduced into such strains (or erythromycin non-producing mutants thereof) to provide the corresponding modified megalomicin compounds in high yields. Those of skill in the art will appreciate that *S. erythraea* contains the desosamine

and mycarose biosynthetic and transfer genes as well as DEBS, which, as noted above, makes the same macrolide aglycone, 6-dEB, as the megalomicin PKS. *S. erythraea* does not make megosamine or its corresponding transferase gene, and does not contain the acylation gene of *Micromonospora megalomicea*. Finally, the  
5 *S. erythraea eryG* gene product converts mycarose to cladinoses, which does not occur in *M. megalomicea*. Thus, the present invention provides a wide variety of *S. erythraea* recombinant host cells, including, for example, those that contain:

- (i) wild-type erythromycin biosynthetic genes with recombinant megosamine biosynthetic and transfer genes, with and without megalomicin  
10 acylation genes;
- (ii) wild-type erythromycin biosynthetic genes except *eryG*, with recombinant megosamine biosynthetic and transfer genes, with and without megalomicin acylation genes; and
- (iii) as in (i) and (ii), except that the *eryA* genes are inactive or deleted and  
15 recombinant *megA* genes have been introduced.

The invention provides other *S. erythraea* strains as well, including those in which any one or more of the erythromycin biosynthetic genes have been deleted or otherwise rendered inactive and in which at least one megalomicin biosynthetic gene has been introduced.

20 For example, the present invention enables one to express the megosamine genes in a *Saccharopolyspora erythraea eryG* mutant in which the erythromycin C made by this mutant is converted to megalomicin A. Alternatively, one could use an erythromycin C high-producing strain of *S. erythraea* in biotransformation methods in which the erythromycin C is fed to a *Streptomyces lividans* strain  
25 carrying only the megosamine biosynthesis and glycosyltransferase genes. As another alternative, one could use a strain of *S. lividans* that carries suitable erythromycin production genes along with the daunosamine biosynthesis genes plus *geneX* and *geneY* of Figure 5, or all of the megosamine biosynthesis genes, to produce megalomicin A.

30 All or some of the megalomicin gene cluster can be easily cloned under control of a suitable promoter in pCK7 or pSET152 either in one or two plasmids and introduced into the *Saccharopolyspora erythraea eryG* mutant. The *actII*-ORF4/*actIp* system and the *phiC31/int* system in pSET function well in this

organism (see Rowe *et al.*, 1998, *Gene*, 216:215-23, incorporated herein by reference). Alternatively, the megosamine biosynthesis genes are introduced into *Streptomyces lividans* on the same plasmids and the production of megalomicin A or its precursor mediated by bioconversion, done by feeding erythronolide B, 3-alpha-mycarosylerythronolide B, erythromycin D or erythromycin C to the *S. lividans* strain.

Lack of adequate resistance to megalomicin A in *S. erythraea* or *S. lividans* is not expected, because both organisms have MLS resistance genes (*ermE* and *mgt/lrm*, respectively), which confer resistance to several 14-membered macrolides (see Cundliffe, 1989, *Annu. Rev. Microbiol.* 43:207-33; Jenkins and Cundliffe, 1991, *Gene* 108:55-62; and Cundliffe, 1992, *Gene*, 115:75-84, each of which is incorporated herein by reference). One can also readily determine the level of resistance of the *S. erythraea eryG* mutant and the *S. lividans* host cells to megalomicin A, both in plate tests and in liquid medium. One can repeat the bioconversion method using an *eryG* mutant of a high erythromycin A producing *S. erythraea* strain (or an *eryB* or *eryC* mutant, as necessary) to determine the level at which megalomicin A can be produced. Furthermore, if experience shows that high level megalomicin A production requires a higher level of resistance to this macrolide than present in *S. erythraea* or *S. lividans*, the necessary megalomicin self-resistance genes will be cloned from *M. megalomicea* and moved into either one of the heterologous hosts. This will be straightforward work since self-resistance genes are usually found in the cluster of macrolide biosynthesis genes and can be identified by their homology to known macrolide resistance genes and(or) by the resistance phenotype they impart to a strain that normally is sensitive.

Alternatively, *geneX* and *geneY* (Figure 5) can be added to cassettes containing the relevant daunosamine (*dnm*) biosynthesis genes (Figure 5) to provide the ability to make TDP-megosamine *in vivo* and attach it to an erythromycin alkycone. The TDP-daunosamine biosynthesis genes can be re-cloned from *Streptomyces peucetius* on two compatible and mutually selectable plasmids. When an *S. lividans* strain containing these two plasmids and the *dnmS* gene for TDP-daunosamine glycosyltransferase is grown in the presence of added epsilon-rhodomyacinone, its glycoside with L-daunosamine, called rhodomycin D,

is produced in good yield. Thus, bioconversion of one of the erythromycins to megalomicin A should be observed when *geneX* and *geneY* are present. One can construct all five combination - the two *N*-dimethyltransferase genes and the three glycosyltransferase genes - to discriminate *geneX* and *geneY* from those connected with mycarose and desosamine biosynthesis and attachment in the megalomicin pathway.

Because the timing of megosamine addition is unknown, one can test erythronolide B, 3- $\alpha$ -mycarosylerythronolide B, erythromycin D and erythromycin C as substrates provided to a strain that expresses the megosamine biosynthetic and transferase genes. There is need to test the C3''' and(or) C4''' acylated metabolites like megalomicin C1, because these metabolites are made from megalomicin A and not the converse, based on the precedents in the biosynthesis of tylosin (see Arisawa *et al.*, 1994, *Appl. Environ. Microbiol.* 60: 2657-2661), carbomycin (see Epp *et al.*, 1989, *Gene* 85:293-301), and midecamycin (see Hara and Hutchinson, 1992, *J. Bacteriol.* 174, 5141-5144). If C-6 glycosylation of erythronolide B or 3- $\alpha$ -mycarosylerythronolide B (Figure 5) happens before addition of desosamine to C-5, then the erythromycin genes might not be able to complete formation of megalomicin A from some mono or diglycoside if the erythromycin glycosyltransferases cannot tolerate a C-6 glycoside. Although unexpected, such an outcome could be circumvented in accordance with the methods of the invention by cloning further megalomicin biosynthesis genes into the appropriate *S. erythraea* background or into *S. lividans* - specifically, the necessary deoxysugar biosynthesis and attachment genes - to create a recombinant strain that produces megalomicin A.

The acyltransferase gene that adds acetate or propionate to the C3''' or C4''' positions of mycarose in megalomicin B, C1 and C2 (Figure 3) is contained within the cosmids of the invention and can be identified by scanning the sequence data for the megalomicin gene cluster to locate homologs of *carE* and *mdmB* or their *acyA* homologs from the tylosin producer. The *carE* and *acyA* genes govern C4''' acylation in the carbomycin and tylosin pathway, respectively. The megalomicin homolog has the equivalent function in megalomicin biosynthesis (but is specific for C3''' and C4''' acylation). The gene can be cloned under control of a suitable promoter and introduced into *S. lividans* to produce the

desired acyl derivative of megalomicin A. Alternatively, introduction of the *carE* gene can form megalomicin B. This gene can be cloned from the carbomycin, spiramycin or tylosin producers.

If the amount of megalomicin produced by an *S. erythraea* or *S. lividans* or  
5 other recombinant host cell is less than desired, yield can be improved by optimizing the growth medium and fermentation conditions, by increasing expression of the gene(s) that appear to be rate limiting, based on the level of pathway intermediates that are accumulated by the recombinant strain constructed, and by reconstructing the *ery*, *dnm*, and megalomicin biosynthesis genes on  
10 vectors like pSET152 that can be integrated into the genome to provide a stabler recombinant strain for strain improvement.

In another embodiment, the present invention provides recombinant vectors encoding one or more of the megosamine, desosamine, and mycarose biosynthetic and transfer genes and heterologous host cells comprising those  
15 vectors. In this embodiment of the invention, the heterologous host cell is typically a cell that is unable to produce the sugar and transfer it to a polyketide unless the vector of the invention is introduced. For example, neither *Streptomyces lividans* nor *S. coelicolor* is naturally capable of making megosamine, desosamine, or mycarose or transferring those moieties to a polyketide. However, the present  
20 invention provides recombinant *Streptomyces lividans* and *S. coelicolor* host cells that are capable of making megosamine, desosamine, and/or mycarose and transferring those moieties to a polyketide.

Moreover, additional recombinant gene products can be expressed in the host cell to improve production of a desired polyketide. As but one non-limiting  
25 example, certain of the recombinant PKS proteins of the invention may produce a polyketide other than or in addition to the predicted polyketide, because the polyketide is cleaved from the PKS by the thioesterase (TE) domain in module 6 prior to processing by other domains on the PKS, in particular, any KR, DH, and/or ER domains in module 6. The production of the predicted polyketide can  
30 be increased in such instances by deleting the TE domain coding sequences from the gene and, optionally, expressing the TE domain as a separate protein. See Gokhale *et al.*, Feb. 1999, "Mechanism and specificity of the terminal thioesterase

domain from the erythromycin polyketide synthase," *Chem. & Biol.* 6: 117-125, incorporated herein by reference.

Thus, in one important aspect, the present invention provides methods, expression vectors, and recombinant host cells that enable the production of megalomicin and hydroxylated and glycosylated derivatives of megalomicin in  
5 heterologous host cells. The present invention also provides methods for making a wide variety of polyketides derived in part from the megalomicin PKS or other biosynthetic genes, as described in the following Section.

#### 10 Section VI: Hybrid PKS Genes

The present invention provides recombinant DNA compounds encoding each of the domains of each of the modules of the megalomicin PKS as well as the other megalomicin biosynthetic enzymes. The availability of these compounds permits their use in recombinant procedures for production of desired portions of  
15 the megalomicin PKS fused to or expressed in conjunction with all or a portion of a heterologous PKS and, optionally, one or more polyketide modification enzymes. These compounds also permit the modification of polyketides with the various megalomicin modification enzymes. The resulting hybrid PKS can then be expressed in a host cell to produce a desired polyketide or modified form thereof.

20 Thus, in accordance with the methods of the invention, a portion of the megalomicin biosynthetic gene coding sequence that encodes a particular activity can be isolated and manipulated, for example, to replace the corresponding region in a different modular PKS gene or modification enzyme gene. In addition, coding sequences for individual proteins, modules, domains, and portions thereof of the  
25 megalomicin PKS can be ligated into suitable expression systems and used to produce the portion of the protein encoded. The resulting protein can be isolated and purified or can may be employed *in situ* to effect polyketide synthesis. Depending on the host for the recombinant production of the domain, module, protein, or combination of proteins, suitable control sequences such as promoters,  
30 termination sequences, enhancers, and the like are ligated to the nucleotide sequence encoding the desired protein in the construction of the expression vector, as described above.



In one important embodiment, the invention thus provides hybrid PKS enzymes and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the invention, a hybrid PKS is a recombinant PKS that comprises all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a first PKS and all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a second PKS. In one preferred embodiment, the first PKS is most but not all of the megalomicin PKS, and the second PKS is only a portion of a non-megalomicin PKS. An illustrative example of such a hybrid PKS includes a megalomicin PKS in which the megalomicin PKS loading module has been replaced with a loading module of another PKS. Another example of such a hybrid PKS is a megalomicin PKS in which the AT domain of extender module 3 is replaced with an AT domain that binds only malonyl CoA. In another preferred embodiment, the first PKS is most but not all of a non-megalomicin PKS, and the second PKS is only a portion of the megalomicin PKS. An illustrative example of such a hybrid PKS includes a rapamycin PKS in which an AT specific for malonyl CoA is replaced with the AT from the megalomicin PKS specific for methylmalonyl CoA. Other illustrative hybrid PKSs of the invention are described below.

Those of skill in the art will recognize that all or part of either the first or second PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See PCT patent application No. WO US99/15047, and Lau *et al.*, *infra*, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. Thus, the desired derivative coding sequences can be synthesized using standard solid phase synthesis methods such as those described by Jaye *et al.*, 1984, *J. Biol. Chem.* 259: 6331, and instruments for automated synthesis are available commercially from, for example, Applied Biosystems, Inc. For purposes of the invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

With this general background regarding hybrid PKSs of the invention, one can better appreciate the benefit provided by the DNA compounds of the invention

that encode the individual domains, modules, and proteins that comprise the megalomicin PKS. As described above, the megalomicin PKS is comprised of a loading module, six extender modules composed of a KS, AT, ACP, and zero, one, two, or three KR, DH, and ER domains, and a thioesterase domain. The DNA compounds of the invention that encode these domains individually or in combination are useful in the construction of the hybrid PKS encoding DNA compounds of the invention. For example, a DNA compound of the invention that encodes an extender module or portion of an extender module is useful in the construction of a coding sequence that encodes a protein subcomponent of a PKS.

10 The DNA compound of the invention that comprises a coding sequence of a PKS subunit protein is useful in the construction of an expression vector that drives expression of the subunit in a host cell that expresses the other subunits and so produces a functional PKS.

The recombinant DNA compounds of the invention that encode the loading module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS loading module is inserted into a DNA compound that comprises the coding sequence for one or more heterologous PKS extender modules. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by that for the coding sequence of the megalomicin PKS loading module provides a novel PKS. Examples include the DEBS, rapamycin, FK-506, FK-520, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS loading module is inserted into a DNA compound that comprises the coding sequence for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

15  
20  
25

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA (propionyl) specific AT with a malonyl CoA (acetyl), ethylmalonyl CoA (butyryl), or other CoA specific AT. In addition, the AT and/or ACP can be replaced by another AT and/or another ACP or an inactivated KS, such as a KS<sup>Q</sup>, an AT, and/or another

30

ACP. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the first  
5 extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS first extender module is inserted into a DNA compound that comprises the coding  
10 sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the megalomicin PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a  
15 sequence that encodes the first extender module of the megalomicin PKS is inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a  
20 hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (which includes inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be  
25 replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a gene for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous first extender module coding sequence can  
30 be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

Those of skill in the art will recognize, however, that deletion of the KR domain of extender module 1 or insertion of a DH domain or DH and KR domains

into extender module 1 will prevent the typical cyclization of the polyketide at the hydroxyl group created by the KR if such hybrid module is employed as a first extender module in a hybrid PKS or is otherwise involved in producing a portion of the polyketide at which cyclization is to occur. Such deletions or insertions can  
5 be useful, however, to create linear molecules or to induce cyclization at another site in the molecule.

As noted above, the invention also provides recombinant PKSs and recombinant DNA compounds and vectors that encode such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are  
10 typically expressed in translational reading frame with the first two extender modules on a single protein, with the remaining modules and domains of a megalomicin, megalomicin derivative, or hybrid PKS expressed as one or more, typically two, proteins to form the multi-protein functional PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the  
15 PKS encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare megalomicin derivative compounds. See U.S. patent application Serial No. 09/492,733, filed 27 Jan. 2000, and PCT publication Nos. WO 00/44717, 99/03986 and 97/02358, each of which is incorporated herein by reference.

20 The recombinant DNA compounds of the invention that encode the second extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS second extender module is inserted into a DNA compound that comprises the  
25 coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that  
30 encodes the second extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; replacing the KR with a KR, a KR and a DH, or a KR, DH, and ER; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the third extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the megalomicin PKS is inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the inactive KR; and/or

replacing the KR with an active KR, or a KR and DH, or a KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module  
5 of the megalomicin PKS, from a gene for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

10 The recombinant DNA compounds of the invention that encode the fourth extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS fourth  
15 extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes  
20 the fourth extender module of the megalomicin PKS is inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a  
25 hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating any one, two, or all three of the ER, DH, and KR; and/or replacing any one, two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and ER. In addition,  
30 the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS (except for the DH and ER domains), from a coding sequence



for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

5           The recombinant DNA compounds of the invention that encode the fifth extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS fifth  
10   extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes  
15   the fifth extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequence for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

          In another embodiment, a portion or all of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to  
20   create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced  
25   with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous fifth extender module coding  
30   sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

          The recombinant DNA compounds of the invention that encode the sixth extender module of the megalomicin PKS and the corresponding polypeptides

encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the sixth extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating the KR or replacing the KR with another KR, a KR and DH, or a KR, DH, and an ER; and/or inserting a DH or a DH and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

The sixth extender module of the megalomicin PKS is followed by a thioesterase domain. This domain is important in the cyclization of the polyketide and its cleavage from the PKS. The present invention provides recombinant DNA compounds that encode hybrid PKS enzymes in which the megalomicin PKS is fused to a heterologous thioesterase or a heterologous PKS is fused to the megalomicin PKS thioesterase. Thus, for example, a thioesterase domain coding sequence from another PKS gene can be inserted at the end of the sixth (or other final) extender module coding sequence in recombinant DNA compounds of the

invention or the megalomicin PKS thioesterase can be similarly fused to a heterologous PKS. Recombinant DNA compounds encoding this thioesterase domain are useful in constructing DNA compounds that encode the megalomicin PKS, a PKS that produces a megalomicin derivative, and a PKS that produces a polyketide other than megalomicin or a megalomicin derivative.

Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:

(i) from fusions of heterologous domain (where heterologous means the domains in a module are derived from at least two different naturally occurring modules) coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS, but also:

(ii) from fusions of heterologous modules (where heterologous module means two modules are adjacent to one another that are not adjacent to one another in naturally occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,

(iii) from expression of one or more megalomicin PKS genes with one or more non-megalomicin PKS genes, including both naturally occurring and recombinant non-megalomicin PKS genes, and

(iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described herein.

An example of a hybrid PKS comprising fused modules results from fusion of the loading module of either the DEBS PKS or the narbonolide PKS (see PCT patent application No. US99/11814, incorporated herein by reference) with extender modules 1 and 2 of the megalomicin PKS to produce a hybrid *megAI* gene. Co-expression of either one of these two hybrid *megAI* genes with the *megAII* and *megAIII* genes in suitable host cells, such as *Streptomyces lividans*, results in expression of a hybrid PKS of the invention that produces 6-deoxyerythronolide B (the polyketide product of the natural *megA* genes) in recombinant host cells. Co-expression of either one of these two hybrid *megAI*

genes with the *eryAII* and *eryAIII* genes similarly results in the production of 6-dEB, while co-expression with the analogous narbonolide PKS genes, *picAII*, *picAIII* and *picAIV*, results in the production of 3-deoxy-3-oxo-6-dEB (3-keto-6-dEB), useful in the production of ketolides, compounds with potent anti-bacterial activity.

Another example of a hybrid PKS comprising a hybrid module is prepared by co-expressing the *megAI* and *megAII* genes with a *megAIII* hybrid gene encoding extender module 5 and the KS and AT of extender module 6 of the megalomicin PKS fused to the ACP of module 6 and the TE of the narbonolide PKS. The resulting hybrid PKS of the invention produces 3-keto-6-dEB. This compound can also be prepared by a recombinant megalomicin derivative PKS of the invention in which the KR domain of module 6 of the megalomicin PKS has been deleted. Moreover, the invention provides hybrid PKSs in which not only the above changes have been made but also the AT domain of module 6 has been replaced with a malonyl-specific AT. These hybrid PKSs produce 2-desmethyl-3-deoxy-3-oxo-6-dEB, a useful intermediate in the preparation of 2-desmethyl ketolides, compounds with potent antibiotic activity.

Another illustrative example of a hybrid PKS includes the hybrid PKS of the invention resulting only from the latter change in the hybrid PKS just described. Thus, co-expression of the *megAI* and *megAII* genes with a hybrid *megAIII* gene in which the AT domain of module 6 has been replaced by a malonyl-specific AT results in the expression of a hybrid PKS that produces 2-desmethyl-6-dEB in recombinant host cells. This compound is a useful intermediate for making 2-desmethyl erythromycins in recombinant host cells of the invention, as well as for making 2-desmethyl semi-synthetic ketolides.

While many of the hybrid PKSs described above are composed primarily of megalomicin PKS proteins, those of skill in the art recognize that the present invention provides many different hybrid PKSs, including those composed of only a small portion of the megalomicin PKS. For example, the present invention provides a hybrid PKS in which a hybrid *eryAI* gene that encodes the megalomicin PKS loading module fused to extender modules 1 and 2 of DEBS is coexpressed with the *eryAII* and *eryAIII* genes. The resulting hybrid PKS produces 6-dEB, the product of the native DEBS. When the construct is expressed in

*Saccharopolyspora erythraea* host cells (either via chromosomal integration in the chromosome or via a vector that encodes the hybrid PKS), the resulting recombinant host cell of the invention produces erythromycins. Another illustrative example is the hybrid PKS of the invention composed of the *megAI* and *eryAII* and *eryAIII* gene products. This construct is also useful in expressing erythromycins in *Saccharopolyspora erythraea* host cells. In a preferred embodiment, the *S. erythraea* host cells are *eryAI* mutants that do not produce 6-deoxyerythronolide B.

Another example is the hybrid PKS of the invention composed of the products of the *picAI* and *picAII* genes (the two proteins that comprise the loading module and extender modules 1 - 4, inclusive, of the narbonolide PKS) and the *megAIII* gene. The resulting hybrid PKS produces the macrolide aglycone 3-hydroxy-narbonolide in *Streptomyces lividans* host cells and the corresponding erythromycins in *Saccharopolyspora erythraea* host cells.

Each of the foregoing hybrid PKS enzymes of the invention, and the hybrid PKS enzymes of the invention generally, can be expressed in a host cell that also expresses a functional *oleP* gene product. The *oleP* gene encodes an oleandomycin modification enzyme, and expression of the gene together with a hybrid PKS of the invention provides the compounds of the invention in which a C-8 hydroxyl, a C-8a or C-8-C-8a epoxide is present.

Recombinant methods for manipulating modular PKS genes to make hybrid PKS enzymes are described in U.S. Patent Nos. 5,672,491; 5,843,718; 5,830,750; and 5,712,146; and in PCT publication Nos. 98/49315 and 97/02358, each of which is incorporated herein by reference. A number of genetic engineering strategies have been used with DEBS to demonstrate that the structures of polyketides can be manipulated to produce novel natural products, primarily analogs of the erythromycins (see the patent publications referenced *supra* and Hutchinson, 1998, *Curr Opin Microbiol.* 1:319-329, and Baltz, 1998, *Trends Microbiol.* 6:76-83, incorporated herein by reference). Because of the similar activity of the megalomicin PKS and DEBS (both PKS enzymes produce the macrolide aglycone 6-dEB), these methods can be readily applied to the recombinant megalomicin PKS genes of the invention.

These techniques include: (i) deletion or insertion of modules to control chain length, (ii) inactivation of reduction/dehydration domains to bypass beta-carbon processing steps, (iii) substitution of AT domains to alter starter and extender units, (iv) addition of reduction/dehydration domains to introduce catalytic activities, and (v) substitution of ketoreductase KR domains to control hydroxyl stereochemistry. In addition, engineered blocked mutants of DEBS have been used for precursor directed biosynthesis of analogs that incorporate synthetically derived starter units. For example, more than 100 novel polyketides were produced by engineering single and combinatorial changes in multiple modules of DEBS. Hybrid PKS enzymes based on DEBS with up to three catalytic domain substitutions were constructed by cassette mutagenesis, in which various DEBS domains were replaced with domains from the rapamycin PKS (see Schweke *et al.*, 1995, *Proc. Nat. Acad. Sci. USA* 92, 7839-7843, incorporated herein by reference) or one more of the DEBS KR domains was deleted. Functional single domain replacements or deletions were combined to generate DEBS enzymes with double and triple catalytic domain substitutions (see McDaniel *et al.*, 1999, *Proc. Nat. Acad. Sci. USA* 96, 1846-1851, incorporated herein by reference). By providing the analogous megalomicin/rapamycin hybrid PKS enzymes, the present invention provides alternative means to make these polyketides.

Methods for generating libraries of polyketides have been greatly improved by cloning PKS genes as a set of three or more mutually selectable plasmids, each carrying a different wild-type or mutant PKS gene, then introducing all possible combinations of the plasmids with wild-type, mutant, and hybrid PKS coding sequences into the same host (see U.S. patent application Serial No. 60/129,731, filed 16 Apr. 1999, and PCT Pub. No. 98/27203, each of which is incorporated herein by reference). This method can also incorporate the use of a KS1° mutant, which by mutational biosynthesis can produce polyketides made from diketide starter units (see Jacobsen *et al.*, 1997, *Science* 277, 367-369, incorporated herein by reference), as well as the use of a truncated gene that leads to 12-membered macrolides or an elongated gene that leads to 16-membered ketolides. Moreover, by utilizing in addition one or more vectors that encode glycosyl biosynthesis and transfer genes, such as those of the present invention for megosamine,



desosamine, oleandrose, cladinose, and/or mycarose (in any combination), a large collection of glycosylated polyketides can be prepared.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant hybrid PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the invention.

#### Avermectin

10 U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemadectin.

15 MacNeil *et al.*, 1992, *Gene 115*: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

#### Candicidin (FR008)

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

#### Epothilone

20 PCT Pub. No. 00/031247 to Kosan.

#### Erythromycin

PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

25 Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.

#### Glycosylation Enzymes

PCT Pub. No. 97/23630 to Abbott.

#### 30 FK-506

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506, *Eur. J. biochem.* 256: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506, *Eur. J. Biochem.* 244: 74-80.

Methyltransferase

- 5 US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520, *J. Bacteriol.* 178: 5243-5248.

10 **FK-520**

PCT Pub. No. 00/20601 to Kosan.

See also Nielsen *et al.*, 1991, *Biochem.* 30:5789-96 (enzymology of pipecolate incorporation).

**Lovastatin**

- 15 U.S. Pat. No. 5,744,350 to Merck.

**Narbomycin (and Picromycin)**

PCT Pub. No. WO US99/61599 to Kosan.

**Nemadectin**

MacNeil *et al.*, 1993, *supra*.

20 **Niddamycin**

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol.* 179: 7515-7522.

**Oleandomycin**

- 25 Swan *et al.*, 1994, Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

PCT Pub. No. 00/026349 to Kosan.

- 30 Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

**Platenolide**

EP Pub. No. 791,656 to Lilly.

### Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

- 5       Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

### Rifamycin

- 10       August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amiclatopsis mediterranei* S669, *Chemistry & Biology*, 5(2): 69-79.

### Soraphen

U.S. Pat. No. 5,716,849 to Novartis.

- 15       Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

### Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

- 20       Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

### Tylosin

EP Pub. No. 791,655 to Lilly.

- 25       Kuhstoss *et al.*, 1996, *Gene* 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

U.S. Pat. No. 5,876,991 to Lilly.

### Tailoring enzymes

Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355.

- 30       Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of PKS genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid PKS-encoding DNA compounds of the invention.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a cognate KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units" *Biochemistry* 38(5):1643-1651, incorporated herein by reference. One can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases", *Science* 284: 482-485, incorporated herein by reference.

The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Even where only two genes are used, there are often two or more modules in the hybrid gene in which all or part of the module is derived from a second (or third) PKS gene. Thus, as one illustrative example, the invention provides a hybrid PKS that contains the naturally occurring loading module and thioesterase domain as well as extender modules one, two, four, and six of the megalomicin PKS and further contains hybrid or heterologous extender modules three and five. Hybrid or heterologous extender modules three and five contain AT domains specific for malonyl CoA and derived from, for example, the rapamycin PKS genes.

The invention also provides libraries of PKS genes, PKS proteins, and ultimately, of polyketides, that are constructed by generating modifications in the megalomicin PKS so that the protein complexes produced have altered activities in one or more respects and thus produce polyketides other than the natural product of the PKS. Novel polyketides may thus be prepared, or polyketides in general prepared more readily, using this method. By providing a large number of

different genes or gene clusters derived from a naturally occurring PKS gene cluster, each of which has been modified in a different way from the native cluster, an effectively combinatorial library of polyketides can be produced as a result of the multiple variations in these activities. As will be further described below, the  
5 metes and bounds of this embodiment of the invention can be described on the polyketide, protein, and the encoding nucleotide sequence levels.

As described above, a modular PKS "derived from" the megalomicin or other naturally occurring PKS includes a modular PKS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilized portion of the  
10 naturally occurring gene. Not all modules need be included in the constructs; however, the constructs can also comprise more than six modules. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS relative to the original (native) PKS. Alteration results when these activities are deleted or are replaced by a different  
15 version of the activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or extender unit, stereochemistry, chain length or cyclization, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a  
20 deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different naturally occurring PKS or from a different region of the megalomicin PKS. Any or all of the megalomicin PKS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of a functional PKS protein is retained in whatever derivative is  
25 constructed. The derivative preferably contains a thioesterase activity from the megalomicin or another PKS.

Thus, a PKS derived from the megalomicin PKS includes a PKS that contains the scaffolding of all or a portion of the megalomicin PKS. The derived PKS also contains at least two extender modules that are functional, preferably  
30 three extender modules, and more preferably four or more extender modules, and most preferably six extender modules. The derived PKS also contains mutations, deletions, insertions, or replacements of one or more of the activities of the functional modules of the megalomicin PKS so that the nature of the resulting

polyketide is altered at both the protein and DNA sequence levels. Particular preferred embodiments include those wherein a KS, AT, or ACP domain has been deleted or replaced by a version of the activity from a different PKS or from another location within the same PKS. Also preferred are derivatives where at least one non-condensation cycle enzymatic activity (KR, DH, or ER) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesized by the PKS.

Conversely, also included within the definition of a PKS derived from the megalomicin PKS are functional non-megalomicin PKS modules or their encoding genes wherein at least one domain or coding sequence therefor of a megalomicin PKS module has been inserted. Exemplary is the use of the megalomicin AT for extender module 2, which accepts a methylmalonyl CoA extender unit rather than malonyl CoA, to replace a malonyl specific AT in another PKS. Other examples include insertion of portions of non-condensation cycle enzymatic activities or other regions of megalomicin synthase activity into a heterologous PKS at both the DNA and protein levels.

Thus, there are at least five degrees of freedom for constructing a hybrid PKS in terms of the polyketide that will be produced. First, the polyketide chain length is determined by the number of extender modules in the PKS, and the present invention includes hybrid PKSs that contain 6, as well as fewer or more than 6, extender modules. Second, the nature of the carbon skeleton of the PKS is determined by the specificities of the acyl transferases that determine the nature of the extender units at each position, e.g., malonyl, methylmalonyl, ethylmalonyl, or other substituted malonyl. Third, the loading module specificity also has an effect on the resulting carbon skeleton of the polyketide. The loading module may use a different starter unit, such as acetyl, butyryl, and the like. As noted above, another method for varying loading module specificity involves inactivating the KS activity in extender module 1 (KS1) and providing alternative substrates, called diketides, that are chemically synthesized analogs of extender module 1 diketide products, for extender module 2. This approach was illustrated in PCT publication Nos. 97/02358 and 99/03986, incorporated herein by reference, wherein the KS1 activity was inactivated through mutation. Fourth, the oxidation state at various positions of the polyketide will be determined by the dehydratase and reductase



portions of the modules. This will determine the presence and location of ketone and alcohol moieties and C-C double bonds or C-C single bonds in the polyketide.

Finally, the stereochemistry of the resulting polyketide is a function of three aspects of the synthase. The first aspect is related to the AT/KS specificity associated with substituted malonyls as extender units, which affects stereochemistry only when the reductive cycle is missing or when it contains only a ketoreductase, as the dehydratase would abolish chirality. Second, the specificity of the ketoreductase may determine the chirality of any beta-OH. Finally, the enoylreductase specificity for substituted malonyls as extender units may influence the stereochemistry when there is a complete KR/DH/ER available.

Thus, the modular PKS systems generally and the megalomicin PKS system particularly permit a wide range of polyketides to be synthesized. As compared to the aromatic PKS systems, the modular PKS systems accept a wider range of starter units, including aliphatic monomers (acetyl, propionyl, butyryl, isovaleryl, and the like.), aromatics (aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and heterocyclics (thiazolyl). Certain modular PKSs have relaxed specificity for their starter units (Kao *et al.*, 1994, *Science, supra*). Modular PKSs also exhibit considerable variety with regard to the choice of extender units in each condensation cycle. The degree of beta-ketoreduction following a condensation reaction can be altered by genetic manipulation (Donadio *et al.*, 1991, *Science, supra*; Donadio *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90: 7119-7123). Likewise, the size of the polyketide product can be varied by designing mutants with the appropriate number of modules (Kao *et al.*, 1994, *J. Am. Chem. Soc.* 116:11612-11613). Lastly, modular PKS enzymes are particularly well known for generating an impressive range of asymmetric centers in their products in a highly controlled manner. The polyketides, antibiotics, and other compounds produced by the methods of the invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it may be beneficial in some instances to generate individual stereoisomers. Thus, the combinatorial potential within modular PKS pathways based on any naturally occurring modular, such as the megalomicin, PKS scaffold is virtually unlimited.

While hybrid PKSs are most often produced by "mixing and matching" portions of PKS coding sequences, mutations in DNA encoding a PKS can also be used to introduce, alter, or delete an activity in the encoded polypeptide. Mutations can be made to the native sequences using conventional techniques. The substrates  
5 for mutation can be an entire cluster of genes or only one or two of them; the substrate for mutation may also be portions of one or more of these genes. Techniques for mutation include preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a PKS subunit using restriction endonuclease digestion. See, e.g., Kunkel, 1985, *Proc.*  
10 *Natl. Acad. Sci. USA* 82: 448; Geisselsoder *et al.*, 1987, *BioTechniques* 5:786. Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) that hybridizes to the native nucleotide sequence, at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within  
15 relatively narrow limits and by keeping the mutant base centrally located. See Zoller and Smith, 1983, *Methods Enzymol.* 100:468. Primer extension is effected using DNA polymerase, the product cloned, and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Identification can be accomplished using the mutant primer as a hybridization  
20 probe. The technique is also applicable for generating multiple point mutations. See, e.g., Dalbie-McFarland *et al.*, 1982, *Proc. Natl. Acad. Sci. USA* 79: 6409. PCR mutagenesis can also be used to effect the desired mutations.

Random mutagenesis of selected portions of the nucleotide sequences encoding enzymatic activities can also be accomplished by several different  
25 techniques known in the art, e.g., by inserting an oligonucleotide linker randomly into a plasmid, by irradiation with X-rays or ultraviolet light, by incorporating incorrect nucleotides during *in vitro* DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants, or by damaging plasmid DNA *in vitro* with chemicals, in accordance with the methods of the present invention.  
30 Chemical mutagens include, for example, sodium bisulfite, nitrous acid, nitrosoguanidine, hydroxylamine, agents which damage or remove bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as 5-bromouracil, 2-aminopurine, or acridine

intercalating agents such as proflavine, acriflavine, quinacrine, and the like.

Generally, plasmid DNA or DNA fragments are treated with chemical mutagens, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

5 In constructing a hybrid PKS of the invention, regions encoding enzymatic activity, i.e., regions encoding corresponding activities from different PKS synthases or from different locations in the same PKS, can be recovered, for example, using PCR techniques with appropriate primers. By "corresponding" activity encoding regions is meant those regions encoding the same general type of activity. For example, a KR activity encoded at one location of a gene cluster  
10 "corresponds" to a KR encoding activity in another location in the gene cluster or in a different gene cluster. Similarly, a complete reductase cycle could be considered corresponding. For example, KR/DH/ER can correspond to a KR alone.

If replacement of a particular target region in a host PKS is to be made,  
15 this replacement can be conducted *in vitro* using suitable restriction enzymes. The replacement can also be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement gene in a donor plasmid and a receptor region in a recipient plasmid. Such systems, advantageously involving plasmids of differing temperature sensitivities are described, for example, in PCT  
20 publication No. WO 96/40968, incorporated herein by reference. The vectors used to perform the various operations to replace the enzymatic activity in the host PKS genes or to support mutations in these regions of the host PKS genes can be chosen to contain control sequences operably linked to the resulting coding sequences in a manner such that expression of the coding sequences can be  
25 effected in an appropriate host.

However, simple cloning vectors may be used as well. If the cloning vectors employed to obtain PKS genes encoding derived PKS lack control sequences for expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This  
30 need not be done individually, but a pool of isolated encoding nucleotide sequences can be inserted into expression vectors, the resulting vectors transformed or transfected into host cells, and the resulting cells plated out into individual colonies. The invention provides a variety of recombinant DNA

compounds in which the various coding sequences for the domains and modules of the megalomicin PKS are flanked by non-naturally occurring restriction enzyme recognition sites.

5 The various PKS nucleotide sequences can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, e.g., a single promoter. The PKS subunit encoding regions can include flanking restriction sites to allow for the easy deletion and insertion of other PKS subunit encoding sequences so that hybrid PKSs can be generated. The design of such unique restriction sites is known to those of skill in the art and can  
10 be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

The expression vectors containing nucleotide sequences encoding a variety of PKS enzymes for the production of different polyketides are then transformed into the appropriate host cells to construct the library. In one straightforward  
15 approach, a mixture of such vectors is transformed into the selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. Each individual colony has the ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some, most, or all of the colonies; the subset of the  
20 transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies are available to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene  
25 cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least four, more typically at least ten, and preferably at least 20, and more preferably at least 50, reflecting similar numbers of different altered PKS gene clusters and PKS gene products. The number of members in the library  
30 is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length enables the production of quite large libraries.

Methods for introducing the recombinant vectors of the invention into suitable hosts are known to those of skill in the art and typically include the use of  $\text{CaCl}_2$  or agents such as other divalent cations, lipofection, DMSO, protoplast transformation, conjugation, infection, transfection, and electroporation. The polyketide producing colonies can be identified and isolated using known techniques and the produced polyketides further characterized. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity.

The libraries of the invention can thus be considered at four levels: (1) a multiplicity of colonies each with a different PKS encoding sequence; (2) the proteins produced from the coding sequences; (3) the polyketides produced from the proteins assembled into a functional PKS; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Of course, combination libraries can also be constructed wherein members of a library derived, for example, from the megalomicin PKS can be considered as a part of the same library as those derived from, for example, the rapamycin PKS or DEBS.

Colonies in the library are induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. The polyketides secreted into the media can be screened for binding to desired targets, such as receptors, signaling proteins, and the like. The supernatants *per se* can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art and can be applied in accordance with the methods of the present invention. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included. Antibiotic activity can be verified using typical screening assays such as those set forth in Lehrer *et al.*, 1991, *J. Immunol. Meth.* 137:167-173, incorporated herein by reference, and in the Examples below.

The invention provides methods for the preparation of a large number of polyketides. These polyketides are useful intermediates in formation of

compounds with antibiotic or other activity through hydroxylation, epoxidation, and glycosylation reactions as described above. In general, the polyketide products of the PKS must be further modified, typically by hydroxylation and glycosylation, to exhibit potent antibiotic activity. Hydroxylation results in the novel polyketides of the invention that contain hydroxyl groups at C-6, which can be accomplished using the hydroxylase encoded by the *eryF* gene, and/or C-12, which can be accomplished using the hydroxylase encoded by the *picK* or *eryK* gene. Also, the *oleP* gene is available in recombinant form, which can be used to express the *oleP* gene product in any host cell. A host cell, such as a *Streptomyces* host cell or a *Saccharopolyspora erythraea* host cell, modified to express the *oleP* gene thus can be used to produce polyketides comprising the C-8-C-8a epoxide present in oleandomycin. Thus the invention provides such modified polyketides. The presence of hydroxyl groups at these positions can enhance the antibiotic activity of the resulting compound relative to its unhydroxylated counterpart.

Methods for glycosylating polyketides are generally known in the art and can be applied in accordance with the methods of the present invention; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means as described herein and in PCT publication No. WO 98/49315, incorporated herein by reference. Preferably, glycosylation with desosamine, mycarose, and/or megosamine is effected in accordance with the methods of the invention in recombinant host cells provided by the invention. In general, the approaches to effecting glycosylation mirror those described above with respect to hydroxylation. The purified enzymes, isolated from native sources or recombinantly produced may be used *in vitro*. Alternatively and as noted, glycosylation may be effected intracellularly using endogenous or recombinantly produced intracellular glycosylases. In addition, synthetic chemical methods may be employed.

The antibiotic modular polyketides may contain any of a number of different sugars, although D-desosamine, or a close analog thereof, is most common. Erythromycin, picromycin, megalomicin, narbomycin, and methymycin contain desosamine. Erythromycin also contains L-cladinose (3-O-methyl mycarose). Tylosin contains mycaminoses (4-hydroxy desosamine), mycarose and



6-deoxy-D-allose. 2-acetyl-1-bromodesosamine has been used as a donor to glycosylate polyketides by Masamune *et al.*, 1975, *J. Am. Chem. Soc.* 97: 3512-3513. Other, apparently more stable donors include glycosyl fluorides, thioglycosides, and trichloroacetimidates; see Woodward *et al.*, 1981, *J. Am. Chem. Soc.* 103: 3215; Martin *et al.*, 1997, *J. Am. Chem. Soc.* 119: 3193; Toshima *et al.*, 1995, *J. Am. Chem. Soc.* 117: 3717; Matsumoto *et al.*, 1988, *Tetrahedron Lett.* 29: 3575. Glycosylation can also be effected using the polyketide aglycones as starting materials and using *Saccharopolyspora erythraea* or *Streptomyces venezuelae* or other host cell to make the conversion, preferably using mutants unable to synthesize macrolides, as discussed in the preceding Section.

Thus, a wide variety of polyketides can be produced by the hybrid PKS enzymes of the invention. These polyketides are useful as antibiotics and as intermediates in the synthesis of other useful compounds, as described in the following section.

#### Section VII: Host Cells Containing Multiple Expression Vectors

A recombinant host cell of the invention may contain nucleic acid encoding a megalomicin PKS domain, module, or protein, or megalomicin modification enzyme at a single genetic locus, *e.g.*, on a single plasmid or at a single chromosomal locus, or at different genetic loci, *e.g.*, on separate plasmids and/or chromosomal loci. By "multiple" is meant two or more; by "vector" is meant a nucleic acid molecule which can be used to transform host systems and which contains an independent expression system containing a coding sequence under control of a promoter and optionally a selectable marker and any other suitable sequences regulating expression. Typical such vectors are plasmids, but other vectors such as phagemids, cosmids, viral vectors and the like can be used according to the nature of the host. Of course, one or more of the separate vectors may integrate into the chromosome of the host (selection may not be required for maintenance of integrated vectors).

In one embodiment, the invention provides a recombinant host cell, which comprises at least two separate autonomously replicating recombinant DNA expression vectors, each of said vectors comprises a recombinant DNA compound encoding a megalomicin PKS domain or a megalomicin modification enzyme

operably linked to a promoter. In another embodiment, the invention provides a recombinant host cell, which comprises at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant  
5 DNA compound encoding a megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter. Preferably, the autonomously replicating recombinant DNA expression vector and/or the modified chromosome further comprises distinct selectable markers.

The above multiple-vector (chromosome) expression systems can also be  
10 used for expressing heterogeneous polyketide biosynthetic enzymes, *e.g.*, for expressing *Micromonospora megalomicea* megalomicin PKS protein, module, or domain or a megalomicin modification enzyme with a PKS protein, module, or domain, or modification enzyme from other origins in the same host cells. By placing various activities on different expression vectors, a high degree of  
15 variation can be achieved in an efficient manner. A variety of hosts can be used; any suitable host cell that can maintain multiple vectors can readily be used. Preferred hosts include *Streptomyces*, yeast, *E. coli*, other actinomycetes, and plant cells, and mammalian or insect cells or other suitable recombinant hosts can also be used. Preferred among yeast strains are *Saccharomyces cerevisiae* and *Pichia*  
20 *pastoris*. Preferred actinomycetes include various strains of *Streptomyces*.

If one chooses to use a host cell that does not naturally produce a polyketide, then one may need to ensure that the recombinant host is modified to also contain a holo ACP synthase activity that effects pantetheinylation of the acyl carrier protein. See PCT Pub. No. WO 97/13845, incorporated herein by  
25 reference. One of the multiple vectors may be used for this purpose. This activation step is necessary for activation of the ACP. The expression system for the holo ACP synthase may be supplied on a vector separate from that carrying a PKS coding sequence or may be supplied on the same vector or may be integrated into the chromosome of the host, or may be supplied as an expression system for a  
30 fusion protein with all or a portion of a polyketide synthase (see U.S. Patent No. 6,033,883, incorporated herein by reference).

It should be noted that in some recombinant hosts, it may also be necessary to activate the polyketides produced through postsynthesis modifications when

polyketides having such modifications are desired. If this is the case for a particular host, the host will be modified, for example by transformation, to contain those enzymes necessary for effecting these modifications. Among such enzymes, for example, are glycosylation enzymes. The use of multiple vectors can  
5 facilitate the introduction of expression systems for such enzymes.

In a preferred embodiment, the multiple vector system is used to assemble rapidly and efficiently a combinatorial library of polyketides and the PKS/modification enzymes that produce them. In an illustrative embodiment, the multiple vector system comprises four different vectors, one comprising the *megAI*  
10 gene, one the *megAII* gene, one the *megAIII* gene, and one the modification enzyme(s) gene(s). Each of these vectors can be modified to make a set of vectors. For example, one set could contain all possible AT substitutions in the loading and first and second extender modules of the *megAI* gene product. Another set could contain expression systems for a variety of different modification enzymes. With  
15 these four vectors sets and by combining each member of each set with each member of the other three sets, a very large library of cells, vector sets, and polyketides can be rapidly and efficiently assembled.

The combinatorial potential of a modular PKS such as the megalomicin PKS (ignoring the additional potential of different modification enzyme systems)  
20 is minimally given by:  $AT_L \times (AT_E \times 4)_M$  where  $AT_L$  is the number of loading acyl transferases,  $AT_E$  is the number of extender acyl transferases, and  $M$  is the number of modules in the gene cluster. The number 4 is present in the formula because this represents the number of ways a keto group can be modified by either  
25 1) no reaction; 2) KR activity alone; 3) KR+DH activity; or 4) KR+DH+ER activity. It has been shown that expression of only the first two modules of the erythromycin PKS resulted in the production of a predicted truncated triketide product (See Kao et al., *J. Am. Chem. Soc.*, 116:11612-11613 ((1994)). A novel  
30 12-membered macrolide similar to methymycin aglycone was produced by expression of modules 1-5 of this PKS in *S. coelicolor* (See Kao et al., *J. Am. Chem. Soc.*, 117:9105-9106 (1995)). This work shows that PKS modules are functionally independent so that lactone ring size can be controlled by the number of modules present.

In addition to controlling the number of modules, the modules can be genetically modified, for example, by the deletion of a ketoreductase domain as described by Donadio et al., *Science*, 252:675-679 (1991); and Donadio et al., *Gene*, 115:97-103 (1992). In addition, the mutation of an enoyl reductase domain  
5 was reported by Donadio, et al., *Proc. Natl. Acad. Sci.*, 90:7119-7123 (1993). These modifications also resulted in modified PKS and thus modified polyketides.

As stated above, in the present invention, the coding sequences for catalytic activities derived from the megalomicin PKS systems found in nature can be used in their native forms or modified by standard mutagenesis techniques to  
10 delete or diminish activity or to introduce an activity into a module in which it was not originally present. For example, a KR activity can be introduced into a module normally lacking that function.

In one embodiment of the invention herein, a single host cell is modified to contain a multiplicity of vectors, each vector contributing a portion of the  
15 synthesis of a megalomicin PKS and modification enzyme (if any) system. Each of the multiple vectors for production of the megalomicin PKS system typically encodes at least two modules, and at least one of the vectors integrates into the chromosome of the host. Integration can be effected using suitable phage or integrating vectors or by homologous recombination. If homologous  
20 recombination is used, the integration event may also be designed to delete endogenous PKS genes residing in the chromosome, as described in the PCT application WO 95/08548. In these embodiments, too, a selectable marker such as hygromycin or thiostrepton resistance can be included in the vector that effects integration.

25 As mentioned above, additional enzymes that effect post-translational modifications to the enzyme systems in the megalomicin PKS may be introduced into the host through suitable recombinant expression systems. In addition, enzymes that activate the polyketides themselves, for example, through glycosylation may be added. It may also be desirable to modify the cell to produce  
30 more of a particular substrate utilized in polyketide biosynthesis. For example, it is generally believed that malonyl CoA levels in yeast are higher than methylmalonyl CoA; if yeast is chosen as a host, it may be desirable to increase

methylmalonyl CoA levels by the addition of one or more biosynthetic enzymes therefor.

The multiple-vector expression system can also be used to make polyketides produced by the addition of synthetic starter units to a PKS that  
5 contains an inactivated ketosynthase (KS) in the first module. As noted above, this modification permits the system to incorporate a suitable diketide thioester such as 3-hydroxy-2-methyl pantonoic acid-N-acetyl cysteamine thioester, or similar thioesters of diketide analogs, as described by Jacobsen et al., *Science*,  
277:367-369 (1997). The construction of PKS modules containing inactivated  
10 ketosynthase regions can be conducted by methods known in the art, such as the method described in U.S. Patent No. 6,080,555 and PCT publication Nos. WO 99/03986 and 97/02358, each of which is incorporated herein by reference, in accordance with the methods of the present invention.

The multiple-vector expression system can be used to produce polyketides  
15 in hosts that normally do not produce them, such as *E. coli* and yeast. It also provides more efficient means to provide a variety of polyketide products by supplying the elements of the introduced PKS, whether in an *E. coli* or yeast host or in other more traditionally used hosts, such as *Streptomyces*. The invention also includes libraries of polyketides prepared using the methods of the invention.

20

#### Section VIII: Compounds

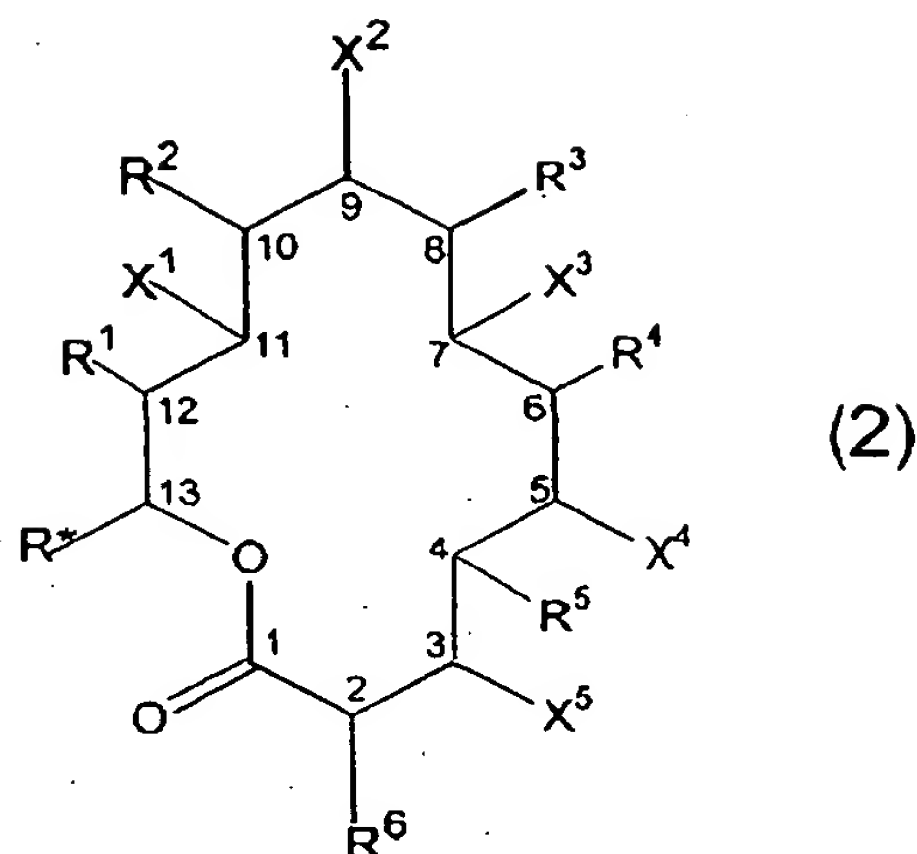
The methods and recombinant DNA compounds of the invention are useful in the production of polyketides. In one important aspect, the invention provides methods for making antibiotic compounds related in structure to erythromycin, a  
25 potent antibiotic compound. The invention also provides novel ketolide compounds, polyketide compounds with potent antibiotic activity of significant interest due to activity against antibiotic resistant strains of bacteria. See Griesgraber *et al.*, 1996, *J. Antibiot.* 49: 465-477, incorporated herein by reference. Most if not all of the ketolides prepared to date are synthesized using  
30 erythromycin A, a derivative of 6-dEB, as an intermediate. In one embodiment, the present invention provides the 3-keto derivatives of the megalomicins for use as antibiotics. In particular, the 3-keto derivative of megalomicin A is a preferred ketolide of the invention. These compounds can be made chemically, substantially

in accordance with the procedures for making ketolides described in the prior art, or in recombinant host cells of the invention in which the megosamine and desosamine biosynthetic and transferase genes are present but which do not make or transfer the mycarose moiety and/or the PKS has been modified to delete the KR domain of extender module 6. The invention also provides methods for making intermediates useful in preparing traditional, 6-dEB- and erythromycin-derived ketolide compounds. See Griesgraber *et al.*, *supra*; Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780; 5,444,051; 5,439,890; 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

As noted above, the hybrid PKS genes of the invention can be expressed in a host cell that contains the desosamine, megosamine, and/or mycarose biosynthetic genes and corresponding transferase genes as well as the required hydroxylase gene(s), which may, for example and without limitation, be either *picK*, *megK*, or *eryK* (for the C-12 position) and/or *megF* or *eryF* (for the C-6 position). The resulting compounds have antibiotic activity but can be further modified, as described in the patent publications referenced above, to yield a desired compound with improved or otherwise desired properties. Alternatively, the aglycone compounds can be produced in the recombinant host cell, and the desired glycosylation and hydroxylation steps carried out *in vitro* or *in vivo*, in the latter case by supplying the converting cell with the aglycone, as described above.

The compounds of the invention are thus optionally glycosylated forms of the polyketide set forth in formula (1) below which are hydroxylated at either the C-6 or the C-12 or both. The compounds of formula (1) can be prepared using the loading and the six extender modules of a modular PKS, modified or prepared in hybrid form as herein described. These polyketides have the formula:





including the glycosylated and isolated stereoisomeric forms thereof;

wherein R\* is a straight chain, branched or cyclic, saturated or unsaturated substituted or unsubstituted hydrocarbyl of 1-15C;

5 each of R<sup>1</sup>-R<sup>6</sup> is independently H or alkyl (1-4C) wherein any alkyl at R<sup>1</sup> may optionally be substituted;

each of X<sup>1</sup>-X<sup>5</sup> is independently two H, H and OH, or =O; or

each of X<sup>1</sup>-X<sup>5</sup> is independently H and the compound of formula (2)

contains a double-bond in the ring adjacent to the position of said X at 2-3, 4-5, 6-

10 7, 8-9 and/or 10-11;

with the proviso that:

at least two of R<sup>1</sup>-R<sup>6</sup> are alkyl (1-4C).

Preferred compounds comprising formula 2 are those wherein at least three of R<sup>1</sup>-R<sup>5</sup> are alkyl (1-4C), preferably methyl or ethyl; more preferably wherein at  
 15 least four of R<sup>1</sup>-R<sup>5</sup> are alkyl (1-4C), preferably methyl or ethyl. Also preferred are those wherein X<sup>2</sup> is two H, =O, or H and OH, and/or X<sup>3</sup> is H, and/or X<sup>1</sup> is OH and/or X<sup>4</sup> is OH and/or X<sup>5</sup> is OH. Also preferred are compounds with variable R\* when R<sup>1</sup>-R<sup>5</sup> is methyl, X<sup>2</sup> is =O, and X<sup>1</sup>, X<sup>4</sup> and X<sup>5</sup> are OH. The glycosylated forms (i.e., mycarose or cladinose at C-3, desosamine at C-5, and/or megosamine  
 20 at C-6) of the foregoing are also preferred.

As described above, there are a wide variety of diverse organisms that can modify compounds such as those described herein to provide compounds with or that can be readily modified to have useful activities. For example, *Saccharopolyspora erythraea* can convert 6-dEB to a variety of useful

compounds. The compounds provided by the present invention can be provided to cultures of *Saccharopolyspora erythraea* and converted to the corresponding derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in the Examples, below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol.* 164(1): 425-433). Also, one can employ other mutant strains, such as *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production. Each of the erythromycins A, B, C, and D has antibiotic activity, although erythromycin A has the highest antibiotic activity. Moreover, each of these compounds can form, under treatment with mild acid, a C-6 to C-9 hemiketal with motilide activity. For formation of hemiketals with motilide activity, erythromycins B, C, and D, are preferred, as the presence of a C-12 hydroxyl allows the formation of an inactive compound that has a hemiketal formed between C-9 and C-12.

Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the compounds of the invention by action of the enzymes endogenous to *Saccharopolyspora erythraea* and mutant strains of *S. erythraea*. Such compounds are useful as antibiotics or as motilides directly or after chemical modification. For use as antibiotics, the compounds of the invention can be used directly without further chemical modification. Erythromycins A, B, C, and D all have antibiotic activity, and the corresponding compounds of the invention that result from the compounds being modified by *Saccharopolyspora erythraea* also have antibiotic activity. These compounds can be chemically modified, however, to provide other compounds of the invention with potent antibiotic activity. For example, alkylation of erythromycin at the C-6 hydroxyl can be used to produce potent antibiotics (clarithromycin is C-6-O-methyl), and other useful modifications are described in, for example, Griesgraber *et al.*, 1996, *J. Antibiot.* 49: 465-477, Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780;

5,444,051; 5,439,890; and 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

For use as motilides, the compounds of the invention can be used directly without further chemical modification. Erythromycin and certain erythromycin  
5 analogs are potent agonists of the motilin receptor that can be used clinically as prokinetic agents to induce phase III of migrating motor complexes, to increase esophageal peristalsis and LES pressure in patients with GERD, to accelerate gastric emptying in patients with gastric paresis, and to stimulate gall bladder contractions in patients after gallstone removal and in diabetics with autonomic  
10 neuropathy. See Peeters, 1999, Motilide Web Site, <http://www.med.kuleuven.ac.be/med/gih/motilid.htm>, and Omura *et al.*, 1987, Macrolides with gastrointestinal motor stimulating activity, *J. Med. Chem.* 30: 1941-3). The corresponding compounds of the invention that result from the compounds of the invention being modified by *Saccharopolyspora erythraea* also have motilide  
15 activity, particularly after conversion, which can also occur *in vivo*, to the C-6 to C-9 hemiketal by treatment with mild acid. Compounds lacking the C-12 hydroxyl are especially preferred for use as motilin agonists. These compounds can also be further chemically modified, however, to provide other compounds of the invention with potent motilide activity.

20 Moreover, and also as noted above, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be  
25 chemically altered after fermentation. In addition to *Saccharopolyspora erythraea*, *Streptomyces venezuelae*, *S. narbonensis*, *S. antibioticus*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans* can also be used. In addition to antibiotic activity, compounds of the invention produced by treatment with *M. megalomicea* enzymes can have antiparasitic activity as well. Thus, the present  
30 invention provides the compounds produced by hydroxylation and glycosylation by action of the enzymes endogenous to *S. erythraea*, *S. venezuelae*, *S. narbonensis*, *S. antibioticus*, *M. megalomicea*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *megAI*, *megAII*, and *megAIII* genes with one or more deletions and/or insertions, including replacements of a *megA* gene fragment with a gene fragment from a heterologous PKS gene, can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Micromonospora megalomicea*, *S. venezuelae*, *S. narbonensis*, *S. antibioticus*, *S. fradiae*, and *S. thermotolerans*.

The compounds of the invention can be produced by growing and fermenting the host cells of the invention under conditions known in the art for the production of other polyketides. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation will contain one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by  
5 reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases caused by infection, a compound of the invention may be administered orally, topically, parenterally, by  
10 inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

Dosage levels of the compounds of the invention are of the order from  
15 about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the invention may be administered on an intermittent  
20 basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral administration to humans may contain from 0.5 mg to 5 gm of  
25 active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention may be formulated within the range of, for example, 0.00001% to 60%  
30 by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight.

It will be understood, however, that the specific dose level for any particular patient will depend on a variety of factors. These factors include the

activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

5 A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the invention and shall not be construed as being a limitation on the scope of the invention or claims.

### Example 1

#### 10 Cloning and Characterization of the Megalomycin Biosynthetic Gene Cluster from *Micromonospora meglomicea*

##### **Experimental Procedures**

##### *Bacterial Strains, Media, and Growth Conditions*

Routine DNA manipulations were performed in *Escherichia coli* XL1 Blue  
15 or *E. coli* XL1 Blue MR (Stratagene) using standard culture conditions (Sambrook  
*et al.*, 1989). *M. meglomicea* subs. *nigra* NRRL3275 was obtained from the  
ATCC collection and cultured according to recommended protocols. For isolation  
of genomic DNA, *M. meglomicea* was grown in TSB (Hopwood *et al.*, 1985) at  
30 °C. *S. lividans* K4-114 (Ziermann and Betlach, 1999), which carries a deletion  
20 of the actinorhodin biosynthetic gene cluster, was used as the host for expression  
of the *megA1-AIII* genes. *S. lividans* strains were maintained on R5 agar at 30°C  
and grown in liquid YEME for preparation of protoplasts (Hopwood *et al.*, 1985).  
*S. erythraea* NRRL2338 was used for expression of the megosamine genes. *S.*  
*erythraea* strains were maintained on R5 agar at 34°C and grown in liquid TSB for  
25 preparation of protoplasts.

##### *Manipulation of DNA and Organisms*

Manipulation and transformation of DNA in *E. coli* was performed by  
standard procedures (Sambrook *et al.*, 1989) or by suppliers protocols. Protoplasts  
30 of *S. lividans* and *S. erythraea* were generated for transformation by plasmid DNA  
using the standard procedure. *S. lividans* transformants were selected on R5 using  
2 ml of a 0.5 mg/ml thiostrepton overlay. *S. erythraea* transformants were selected  
on R5 using 1.5 ml of a 0.6 mg/ml apramycin overlay.



*Isolation of the meg gene cluster*

A cosmid library was prepared in SuperCos (Stratagene) from *M. megalomicea* total DNA partially digested with *Sau3A* I, and introduced into *E. coli* using a Gigapack III XL (Stratagene) *in-vitro* packaging kit. <sup>32</sup>P-labelled DNA probes encompassing the KS2 domain from *ery* DEBS, or a mixture of segments encompassing modules 1 and 2 from *ery* DEBS were used separately to screen the cosmid library by colony hybridization. Several colonies which hybridized with the probes were further analyzed by sequencing the ends of their cosmid inserts using T3 and T7 primers. BLAST (Altschul *et al.*, 1990) analysis of the sequences revealed several colonies with DNA sequences highly homologous to genes from the *ery* cluster. Together with restriction analysis, this led to the isolation of two overlapping cosmids, pKOS079-93A and pKOS079-93D which covered ~45 kb of the *meg* cluster. A 400 bp PCR fragment was generated from the left end of and pKOS079-93D and used to reprobe the cosmid library. Likewise, a 200 bp PCR fragment generated from the right end of pKOS079-93A was used to reprobe the cosmid library. Analysis of hybridizing colonies as described above resulted in identification of two additional cosmids, pKOS079-138B and pKOS79-124B which overlap the previous two cosmids. BLAST analysis of the far left and right end sequences of these cosmids indicated no homology to any known genes related to polyketide biosynthesis and therefore indicates that the set of four cosmids spans the entire megalomicin biosynthetic gene cluster.

*DNA sequencing and analysis*

PCR-based double stranded DNA sequencing was performed on a Beckman CEQ 2000 capillary sequencer using reagents and protocols provided by the manufacturer. A shotgun library of the entire cosmid pKOS079-93D insert was made as follows: DNA was first digested with *Dra* I to eliminate the vector fragment, then partially digested with *Sau3A* I. After agarose electrophoresis, bands between 1-3 kb were excised from the gel and ligated with *Bam*H I digested pUC19. Another shotgun library was generated from a 12 kb *Xho* I/*Eco*R I fragment subcloned from cosmid pKOS079-93A to extend the sequence to the *megF* gene. A 4 kb *Bgl* II/*Xho* I fragment from cosmid pKOS079-138B was

sequenced by primer walking to extend the sequencing to the *megT* gene.

Sequence was assembled using Sequencher (Gene Codes Corp.) software package and analyzed with MacVector (Oxford Molecular Group) and the NCBI BLAST server ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)).

5

### Plasmids

Plasmid pKOS108-6 is a modified version of pKAO127'kan' (Ziermann and Betlach, 1999; Ziermann and Betlach, 2000) in which the *eryAI*-III genes between the *Pac* I and *EcoR* I sites have been replaced with the *megAI*-III genes.

10 This was done by first substituting a synthetic nucleotide DNA duplex (5'-TAAGAATTCGGAGATCTGGCCTCAGCTCTAGAC (SEQ ID NO: 21), complementary oligo 5'-

AATTGTCTAGAGCTGAGGCCAGATCTCCGAATTCTTAAT (SEQ ID NO: 22)) between the *Pac* I and *EcoR* I sites of the pKAO127'kan' vector fragment.

15 The 22 kb *EcoR* I/*Bgl* II fragment from cosmid pKOS079-93D containing the *megAI*-II genes was inserted into *EcoR* I and *Bgl* II sites of the resulting plasmid to generate pKOS024-84. A 12 kb *Bgl* II/*BbvC* I fragment containing the *megAIII* and part of the *megCII* gene was subcloned from pKOS079-93A and excised as a *Bgl* II/*Xba* I fragment and ligated into the corresponding sites of pKOS024-84 to  
20 yield the final expression plasmid pKOS108-06.

The megosamine integrating vector, pKOS97-42, was constructed as follows: A subclone was generated containing the 4 kb *Xho* I/*Sca* I fragment from pKOS79-138B together with the 1.7 kb *Sca* I/*Pst* I fragment from pKOS79-93D in Litmus 28 (Stratagene). The entire 5.7 kb fragment was then excised as a *Spe* I/*Pst* I fragment and combined with the 6.3 kb *Pst* I/*EcoR* I fragment from KOS79-93D and *EcoR* I/*Xba* I digested pSET152 (Bierman *et al.*, 1992) to construct plasmid pKOS97-42.

### Production and analysis of secondary metabolites

30 Fermentation for production of polyketide, LC/MS analysis, and quantification of 6-dEB for *S. lividans* K4-114/pKOS108-6 and *S. lividans* K4-114/pKAO127'kan' were essentially as previously described (Xue *et al.*, 1999). *S. erythraea* NRRL2338 and *S. erythraea*/pKOS97-42 were grown for 6 days in F1

media (Brünker *et al.*, 1998). Samples of broth were clarified in a microcentrifuge (5 min, 13,000 rpm). For LC/MS preparation, isopropanol was added to the supernatant (1:2 ratio) and centrifuged again. Erythromycins and megalomicins were detected by electrospray mass spectrometry and quantity was determined by  
5 evaporative light scattering detection (ELSD). The LC retention time and mass spectra of erythromycin and megalomicins were identical to known standards.

#### *Nucleotide sequence of the meg gene cluster*

A series of 4 overlapping inserts containing the *meg* cluster (Figure 9) were  
10 isolated from a cosmid library prepared from total genomic DNA of *M. megalomicea* and covers > 100 kb of the genome. A contiguous 48 kb segment which encodes the megalomicin PKS and several deoxysugar biosynthetic genes was sequenced and analyzed. The segment contains 17 complete ORFs as well as an incomplete ORF at each end, organized as shown in Figure 9.

15 *PKS genes.* The ORFs *megAI*, *megAII* and *megAIII* encode the polyketide synthase responsible for synthesis of 6-dEB. The enzyme complex, *meg* DEBS, is highly similar to *ery* DEBS, with each of the three predicted polypeptides sharing an average of 83% overall similarity with their *ery* PKS counterpart. Both PKSs are composed of 6 modules (2 modules per polypeptide) and each module is  
20 organized in the identical manner (Figure 9). A dendrogram analysis (Schwecke *et al.*, 1995) employing 70 acyltransferase (AT) domains revealed that the 6 *meg* extender AT domains cluster with AT domains that incorporate methylmalonyl CoA (not shown). The loading module of *meg* DEBS also lacks a KS<sup>Q</sup> domain which is utilized by most macrolide PKSs for decarboxylation of the starter unit to  
25 initiate polyketide synthesis (Bisang *et al.*, 1999; Kuhstoss *et al.*, 1996; Kakavas *et al.*, 1997; Xue *et al.*, 1998), implying that priming begins with a propionate unit. In addition, a conserved Gly to Pro substitution in the NADPH-binding region of the ketoreductase (KR) domain of module 3 is observed in *meg* DEBS, which has been proposed to account for its inactivity in *ery* DEBS (Donadio *et al.*, 1991).

30 *Deoxysugar genes.* BLAST (Altschul *et al.*, 1990) analysis of the genes flanking the PKS indicated that 12 complete ORFs and 1 partial ORF appear to encode functions required for synthesis of one of the three megalomicin deoxysugars. Assignment of each ORF to a specific deoxysugar pathway was

made based on comparison to the *ery* genes and other related genes involved in deoxysugar biosynthesis (Table 2).

Table 2. Deduced functions of genes identified in the megalomicin gene cluster.

<i>Gene</i>	<i>Closest Match</i> (polypeptide) <sup>a</sup>	<i>% Sim<sup>a</sup></i>	<i>Proposed</i> <i>Pathway</i>	<i>Proposed Function</i>	<i>Reference</i>
<i>megT</i>	EryBVI		Mycarose/ Megosamine	2,3-Dehydratase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megDVI</i>	EryCII	63	Megosamine	3,4-Isomerase	(Summers <i>et al.</i> , 1997)
<i>megDI</i>	EryCIII	79	Megosamine	Glycosyltransferase	(Summers <i>et al.</i> , 1997)
<i>megY</i>	AcyA ( <i>S.</i> <i>thermotolerans</i> )	52		Mycarose O-acyl- transferase	(Arisawa <i>et al.</i> , 1994)
<i>megDII</i>	EryCI	58	Megosamine	Aminotransferase	(Dhillon <i>et al.</i> , 1989; Summers <i>et al.</i> , 1997)
<i>megDIII</i>	DesVI ( <i>S.</i> <i>venezuelae</i> )	61	Megosamine	Dimethyltransferase	(Xue <i>et al.</i> , 1998)
<i>megDIV</i>	DmnU ( <i>S.</i> <i>peucetius</i> )	65	Megosamine	3,5-Epimerase	(Olano <i>et al.</i> , 1999)
<i>megDV</i>	Dehydrogenase ( <i>A. orientalis</i> )	61	Megosamine	4-Ketoreductase	(Summers <i>et al.</i> , 1997; van Wageningen <i>et al.</i> , 1998)
<i>megDVII</i>	EryBII	73	Megosamine	2,3-Reductase	(Summers <i>et al.</i> , 1997)
<i>megBV</i>	EryBV	86	Mycarose	Glycosyltransferase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megBIV</i>	EryBIV	80	Mycarose	4-Ketoreductase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megAI</i>	EryAI	81	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megAII</i>	EryAII	85	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megAIII</i>	EryAIII	83	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megCII</i>	EryCII	82	Desosamine	3,4-Isomerase	(Summers <i>et al.</i> , 1997)
<i>meg CIII</i>	EryCIII	89	Desosamine	Glycosyltransferase	(Summers <i>et al.</i> , 1997)
<i>megBII</i>	EryBII	87	Mycarose	2,3-Reductase	(Summers <i>et al.</i> , 1997)
<i>megH</i>	EryH	84		Thioesterase	(Haydock <i>et al.</i> , 1991)
<i>megF</i>	EryF			C-6 Hydroxylase	(Weber <i>et al.</i> , 1991)

5 a. Determined by BLASTX analysis using default parameters.

Three ORFs, *megBV*, *megCIII* and *megDI*, encode glycosyltransferases, apparently one for attachment of each deoxysugar to the macrolide. MegBV was most similar to EryBV, the erythromycin mycarosyltransferase, and hence was assigned to the mycarose pathway in the *meg* cluster. The closest match for both of the remaining glycosyltransferases was EryCIII, the desosaminytransferase in erythromycin biosynthesis. Given the higher degree of similarity between EryCIII and MegCIII (Table 2), MegCIII was designated the desosaminytransferase, leaving MegDI as the proposed megosaminytransferase. In similar fashion, assignments were made accordingly for; MegCII and MegDVI, two putative 3,4-isomerases similar to EryCII; MegBII and MegDVII, 2,3-reductases homologous to EryBII; MegBIV and MegDV, putative 4-ketoreductases similar to EryBIV (Table 2). The remaining ORFs involved in deoxysugar biosynthesis, *megT*, *megDII*, *megDIII* and *megDIV*, each encode a putative 2,3-dehydratase, aminotransferase, dimethyltransferase and 3,5-epimerase, respectively (Table 2). Since both the megosamine and desosamine pathways require an aminotransferase and a dimethyltransferase, and since mycarose and megosamine each require a 2,3-dehydratase and a 3,5-epimerase, assignments of these four genes to a specific pathway could not be made on the basis of sequence comparison alone. However, the latter three are implicated in megosamine biosynthesis by experiments described below.

*Other genes.* Two additional complete ORFs, designated *megY* and *megH* and an incomplete ORF, designated *megF*, were also identified in the cluster. MegH and MegF share high degrees of similarity with EryH and EryF. EryH and homologs in other macrolide gene clusters are thioesterase-like proteins with unknown function in polyketide gene clusters (Haydock *et al.*, 1991; Xue *et al.*, 1998; Butler *et al.*, 1999; Tang *et al.*, 1999). EryF encodes the erythronolide B C-6 hydroxylase (Figure 8) (Weber *et al.*, 1991; Andersen and Hutchinson, 1992). MegY does not have an *ery* counterpart but appears to belong to a (small) family of *O*-acyltransferases that transfer short acyl chains to macrolides. Two classes exist: AcyA and MdmB transfer acetyl or propionyl groups to the C-3 hydroxyls on 16-membered macrolide rings (Arisawa *et al.*, 1994; Hara and Hutchinson, 1992); CarE and Mpt transfer isovalerate or propionate to the mycarosyl moiety of carbomycin and midecamycin, respectively (Epp *et al.*, 1989; Arisawa *et al.*, 1993;

Gu *et al.*, 1996). The structures of various megalomicins suggest that MegY belongs to the latter class and is the acyltransferase which converts megalomicin A to megalomicins B, C1, or C2 (verified experimentally below).

5 *Heterologous expression of the meg PKS genes.*

The wild type and genetically modified versions of the *ery* DEBS have been used extensively in heterologous *Streptomyces* hosts for enzyme studies and the production of novel polyketide compounds. Given the similarities between the *ery* and *meg* DEBSs, production characteristics were compared in a commonly  
10 used *Streptomyces* host strain. The three *megA* ORFs were cloned into the expression plasmid pKAO127'kan' (Ziermann and Betlach, 1999) in place of the *eryA* ORFs. Both plasmids, pKAO127'kan' encoding *ery* DEBS and pKOS108-06 encoding *meg* DEBS, were introduced in *Streptomyces lividans* K4-114 and the production of 6-dEB was determined in shake-flask fermentations. The production  
15 profiles were similar in both cases and the maximum titer of 6-dEB was between 30-40 mg/L. In addition, both PKSs produced small amounts (~5%) of 8,8a-deoxyoleandolide, which results from the priming of the PKS with acetate instead of propionate (Kao *et al.*, 1994b). This observation indicates that the loading AT domains of the PKSs display similar relaxed specificities towards starter units.

20

*Conversion of erythromycin to megalomicin in S. erythraea.*

An examination of the *meg* cluster revealed that the putative megosamine biosynthetic genes are clustered directly upstream of the PKS genes. If the hypothesis that these genes are sufficient for biosynthesis and attachment of  
25 megosamine to an erythromycin intermediate is correct, then functional expression of these genes in a strain which produces erythromycin, such as *S. erythraea*, should result in production of megalomicin. A 12 kb DNA fragment carrying all the genes between the leftmost *Xho*I site and the *Eco*RI site (Figure 9) was integrated in the chromosome of *S. erythraea* using the site-specific integrating  
30 vector pSET152 (Bierman *et al.*, 1992). It was surmised that the left and right ends of this fragment would contain necessary promoter regions for transcription of the convergent set of genes in *M. megalomicea* and that they would likely operate in *S. erythraea*.



Fermentation broth from *S. erythraea*/KOS97-42, which contains the integrated *meg* genes, was analyzed by LC/MS and compared to LC/MS profiles of the parent *S. erythraea* strain without the *meg* genes, as well as to megalomicin standards purified from *M. megalomicea*. The new strain was found to produce a mixture of erythromycin A and various megalomicins (~4:1 ratio), thereby showing that the predicted megosamine biosynthetic and glycosyltransferase genes are contained within the cloned *meg* fragment. The two most abundant congeners identified were megalomicins B and C1. Megalomicin A and C2 were also detected in smaller amounts. The presence of the megalomicins B, C1 and C2 also provides direct evidence for the function of the *O*-acyl transferase, MegY, which is present in the integrated *meg* fragment.

### Discussion

The homologies observed among modular PKSs enabled the use of *ery* PKS genes to clone the *meg* biosynthetic gene cluster from *M. megalomicea*. The close similarities between the megalomicin and erythromycin biosynthetic pathways is also reflected in the overall organization of their genes and in the high degree of homology of the corresponding individual gene-encoded polypeptides. Production of 6-dEB from *meg* DEBS in *S. lividans* and conversion of erythromycin to megalomicin using the *megD* genes in *S. erythraea* provides direct evidence that the identified gene cluster is responsible for synthesis of megalomicin.

As seen in Figure 9, the ~ 40 kb segments of the two clusters beginning with *ery/megBV* on the left through the *ery/megF* genes retain a nearly identical organizational arrangement. The notable differences in this region are *eryG* and IS1136 which are absent from the segment of the *meg* cluster analyzed. The *eryG* gene encodes an S-adenosylmethionine (SAM)-dependent mycarosyl methyltransferase that converts erythromycin C to erythromycin A (Figure 8) (Weber *et al.*, 1990; Haydock *et al.*, 1991). The mycarose moiety is modified by esterification (MegY) in megalomicin biosynthesis (Figure 8) and, therefore, the absence of an *eryG* homolog would be expected in the *meg* cluster. The IS1136 element located between *eryAI* and *eryAII* (Donadio and Staver, 1993) is not

known to play a role in erythromycin biosynthesis and its origin in the *ery* cluster has not been determined.

Upstream of the common *meg/eryBIV* and *BV* genes, the gene clusters diverge. The ~ 6 kb segment between *eryBV* and *eryK*, the left border of the *ery* gene cluster (Pereda *et al.*, 1997), contains the remaining genes required for mycarose (*eryBVI* and *BVII*) and desosamine biosynthesis (*eryCIV*, *CV*, and *CVI*) and the C-12 hydroxylase (*eryK*) (Stassi *et al.*, 1993). In contrast, the region upstream of *megBV* encodes a set of genes (*megDI-DVII* and *megY*) which can account for all the activities unique to megalomicin biosynthesis (Figure 9). Since introduction of this *meg* DNA segment into *S. erythraea* results in production of megalomicins, it is clear that these genes encode the functions for TDP-megosamine biosynthesis and transfer to its putative substrate erythromycin C, and to acylate megalomicin A (Figure 8). The remaining region upstream of *megDVI* should therefore encode genes only for mycarose and desosamine biosynthesis.

Olano *et al.* (Olano *et al.*, 1999) have recently described a pathway for biosynthesis of TDP-L-daunosamine, a deoxysugar component of the antitumor compounds daunorubicin and doxorubicin produced by *Streptomyces peucetius*. Their pathway proposes four steps from the intermediate TDP-4-keto-6-deoxyglucose controlled by the gene cluster *dnmJQTUVZ*, although the functions for *dnmQ* and *dnmZ* could not be identified and the precise order of reactions in the pathway could not be determined. The genes *dnmT*, *dnmU*, *dnmJ* and *dnmV* each have proposed counterparts in the *meg* cluster, *megT*, *megDIV*, *megDII*, and *megDV*, respectively (see Figure 10)

It is possible to describe a pathway to convert TDP-2,6-dideoxy-3,4-diketo-D-hexose (or its enol tautomer), the last intermediate common to the mycarose and megosamine pathways, to TDP-megosamine through the sequence of 5-epimerization, 4-ketoreduction, 3-amination, and 3-*N*-dimethylation employing the genes *megDIV*, *megDV*, *megDII*, and *megDIII*. This employs the same functions proposed for biosynthesis of TDP-daunosamine by Olano *et al.*, but in a different sequential order. However, it does not account for the *megDVI* and *megDVII* genes since their activities are not required for this route. A parallel pathway which employs these genes is also shown in Figure 10. In this alternate route, 2,3-reduction and 3,4-tautomerization are performed by the *megDVII* and

*megDVI* gene products, respectively. A unified single pathway that employs both 4-ketoreduction (*megDV*) and 2,3-reduction (*megDVII*) could not be determined. Because the entire gene set from *megDVI* through *megDVII* was introduced in *S. erythraea* to produce TDP-megosamine, it is not possible to determine which, if  
5 either, of the two alternative pathways is operative, but this can be addressed through systematic gene disruption and complementation.

The 48 kb segment sequenced also contains genes required for synthesis of TDP-L-mycarose and TDP-D-desosamine (Fig 10). For the latter, *megCII*, which encodes a putative 3,4-isomerase, the first step in the committed TDP-desosamine  
10 pathway, appears to be translationally coupled to *megAIII*, almost exactly as its erythromycin counterpart, *eryCII*, was found translationally coupled to *eryAIII* (Summers *et al.*, 1997). The high degree of similarity between MegCII and EryCII suggests that the pathway to desosamine in the megalomicin- and erythromycin-producing organisms are most likely the same. Similarly, the finding that *megBII*  
15 and *megBIV*, encoding a 2,3-reductase and 4-ketoreductase, contain close homologs in the mycarose pathway for erythromycin also suggests that TDP-L-mycarose synthesis in the two host organisms is the same.

Of interest are the two genes that encode putative 2,3-reductases, *megBII* and *megDVII*. Because MegBII most closely resembles EryBII, a known mycarose  
20 biosynthetic enzyme (Weber *et al.*, 1990), and because *megBII* resides in the same location of the *meg* cluster as its counterpart in the *ery* cluster, *megBII* is assigned to the mycarose pathway and *megDVII* to the megosamine pathway. Furthermore, the lower degree of similarity between MegDVII and either EryBII or MegBII (Table 2) provides a basis for assigning the opposite L and D isomeric substrates  
25 to each of the enzymes (Figure 10). Finally, *megT*, which encodes a putative 2,3-dehydratase, is also related to a gene in the *ery* mycarose pathway, *eryBVI*. In *S. erythraea*, the proposed intermediate generated by EryBVI represents the first committed step in the biosynthesis of mycarose (Figure 10). However, the proposed pathways in Figure 10 suggest this may be an intermediate common to  
30 both mycarose and megosamine biosynthesis in *M. megalomicea*. Therefore, *megT* is named following the designation of the equivalent gene in the daunosamine pathway, *dnmT* (Olano *et al.*, 1999)

The preferred host-vector system for expression of *meg* DEBS described here has been used previously for the heterologous expression of modular PKS genes from the erythromycin (Kao *et al.*, 1994a; Ziermann and Betlach, 1999), picromycin (Tang *et al.*, 1999) and oleandomycin pathways, as well as for the generation of novel polyketide backbones where domains have been removed, added or exchanged in various combinations (McDaniel *et al.*, 1999). Recently, hybrid polyketides have been generated through the co-expression of subunits from different PKS systems (Tang *et al.*, 2000).

Expression of the *megDVI-megDVII* segment in *S. erythraea* and the corresponding production of megalomicins in this host establishes the likely order of sugar attachment in megalomicin synthesis. Furthermore, it provides a means to produce megalomicin in a more genetically friendly host organism, leading to the creation of megalomicin analogs by manipulating the PKS. Over 60 6-dEB analogs have been produced by combinatorial biosynthesis using the *ery* PKS (McDaniel *et al.*, 1999; Xue *et al.*, 1999). The titers of megalomicin could also be significantly increased above the 5 mg/L obtained from *M. megalomycina* by introducing the genes into an industrially optimized strain of *S. erythraea*, many of which can produce as much as 10 g/L of erythromycin.

## References

- Kao, C.M., Katz, L. and Khosla, C. (1994a) Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265**: 509-512.
- Kao, C.M., Luo, G., Katz, L., Cane, D.E. and Khosla, C. (1994b) Engineered biosynthesis of a triketide lactone from an incomplete modular polyketide synthase. *J. Am. Chem. Soc.* **116**: 11612-11613.
- McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M. *et al.* (1999) Multiple genetic modifications of the erythromycin gene cluster to produce a library of novel "unnatural" natural products. *Proc. Natl. Acad. Sci. USA* **96**: 1846-1851.
- Olano, C., Lomovskaya, N., Fonstein, L., Roll, J.T. and Hutchinson, C.R. (1999) A two-plasmid system for the glycosylation of polyketide antibiotics:

bioconversion of e-rhodomyacinone to rhodomyacin D. *Chem. & Biol.* 6: 845-855.

- 5 Tang, L., Fu, H., Betlach, M.C. and McDaniel, R. (1999) Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. *Chem. & Biol.* 6: 553-558.
- Tang, L., Fu, H. and McDaniel, R. (2000) Formation of functional heterologous complexes using subunits from the picromycin, erythromycin, and oleandomycin polyketide synthases. *Chem. & Biol.* 7: 77-84.
- 10 Weber, J.M., Leung, J.O., Maine, G.T., Potenz, R.H., Paulus, T.J. and DeWitt, J.P. (1990) Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J. Bacteriol.* 172: 2372-2383.
- Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. and McAlpine, J.B. (1991) An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* 252: 114-117.
- 15 Xue, Q., Ashley, G., Hutchinson, C.R. and Santi, D.V. (1999) A multi-plasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. USA* 96: 11740-11745.
- Xue, Y., Zhao, L., Liu, H.-w. and Sherman, D.H. (1998) A gene cluster for the macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* 95: 12111-12116.
- 20 Ziermann, R. and Betlach, M. (2000) A two-vector system for the production of recombinant polyketides in *Streptomyces*. *J. Ind. Microbiol. Biotech.* 24: 46-50.
- Ziermann, R. and Betlach, M.C. (1999) Recombinant polyketide synthesis in *Streptomyces*: Engineering of improved host strains. *Biotechniques* 26: 106-110.
- 25

## Example 2

### Stabilizing meg PKS Expression Plasmid by Codon Engineering

#### 30 *Materials and methods*

All bacterial strains were cultured and transformed as described in Example 1.

### *Fermentation of Streptomyces and diketide feeding*

Primary *Streptomyces* transformants were picked and placed in 6 mL of TSB liquid medium with 50 µg/L of thiostrepton and grown at 30°C. When the culture showed some growth (3-4 days), it was transferred into a 250 mL flask containing 50 mL of R6 medium (pH 7.0) with 25 µg/L of thiostrepton and 1g/L of diketide ((2s,3R)2-methyl-3-hydroxyhexanoate N-propionyl cysteamine thioester) and placed in a 30°C incubator for 7 days.

### 10 *Changing codons and making plasmids*

There are several identical sequences in the coding sequences for module 2 and module 6 of the megalomicin PKS gene cluster. Expression plasmids containing the full length megalomicin PKS appeared to be somewhat unstable and subject to deletion in *recA*<sup>+</sup> strains like ET124567 and *Streptomyces* by intra-plasmid homologous recombination. To prevent significant homologous recombination and so stabilize expression plasmids, the codons of two regions of the module 6 coding sequence that are identical to regions in the module 2 coding sequence were changed without changing the sequence of protein encoded. The two regions changed in module 6 were from the 26739<sup>th</sup> base to 27,267<sup>th</sup> base and from position 27,697<sup>th</sup> base to 27,987<sup>th</sup> base, which were identical to the region from position 6810<sup>th</sup> base to 7338<sup>th</sup> base and regions from position 7778<sup>th</sup> base to 8068<sup>th</sup> base, respectively. The start codon of the loading domain of the meg PKS was set to be the 1<sup>st</sup> base. These sequences are shown below

```

25 > 6810-7338 Sequence in Module 2
   TTGCAGCGGTTGTCGGTGGCGGTGCGGGAGGGGCGTGGGTGTTGGGTGTGGTGGTGGGT
   TCGGCGGTGAATCAGGATGGGGCGAGTAATGGGTGGCGGCGCCGTCGGGGGTGGCGCAG
   CAGCGGGTGATTTCGGCGGGCGTGGGGTTCGTGCGGGTGTGTCGGGTGGGGATGTGGGTGTG
   GTGGAGGCGCATGGGACGGGGACGCGGTTGGGGGATCCGGTGGAGTTGGGGGCGTTGTTG
30 GGGACGTATGGGGTGGGTTCGGGGTGGGGTGGGTCCGGTGGTGGTGGGTTCGGTGAAGGCG
   AATGTGGGTCATGTGCAGGCGGCGGCGGGTGTGGTGGGTGTGATCAAGGTGGTGTGGGG
   TTGGGTTCGGGGGTTGGTGGGTCCGATGGTGTGTCGGGGTGGGTTCGGGGGTTGGTGGAT
   TGGTCGTTCGGGTGGGTGGTGGTGGCGGATGGGGTGCGGGGGTGGCCGGTGGGTGTGGAT
   GGGGTGCGTTCGGGGTGGGGTGTTCGGCGTTTGGGGTGTTCGGGGACGAAT (SEQ ID NO: 23)
35 > 26736-27267 Sequence in Module 6
   CTGCAGCGGTTGTCGGTGGCGGTGCGGGAGGGGCGTGGGTGTTGGGTGTGGTGGTGGGT
   TCGGCGGTGAATCAGGATGGGGCGAGTAATGGGTGGCGGCGCCGTCGGGGGTGGCGCAG
   CAGCGGGTGATTTCGGCGGGCGTGGGGTTCGTGCGGGTGTGTCGGGTGGGGATGTGGGTGTG
   GTGGAGGCGCATGGGACGGGGACGCGGTTGGGGGATCCGGTGGAGTTGGGGGCGTTGTTG
40 GGGACGTATGGGGTGGGTTCGGGGTGGGGTGGGTCCGGTGGTGGTGGGTTCGGTGAAGGCG
   AATGTGGGTCATGTGCAGGCGGCGGCGGGTGTGGTGGGTGTGATCAAGGTGGTGTGGGG

```



TTGGGTCGGGGGTTGGTGGGTCCGATGGTGTGTCGGGGTGGGTTGTCGGGGTGGTGGAT  
 TGGTCGTCGGGTGGGTGGTGGCGGATGGGGTGCGGGGGTGGCCGGTGGGTGTGGAT  
 GGGGTGCGTCGGGGTGGGTGTCGGCGTTTGGGGTGTGGGGACGAAT (SEQ ID NO: 24)  
 > 26736-27267 Sequence with Codon Changes  
 5 CTGCAGCGCCTCTCCGTGCGCGTCCGCGAGGGCCGCGAGTCCTCGGCGTCGTGTCGGC  
 TCGGCCGTCAACCAAGACGGCGCGTCAAACGGCCTCGCCGCGCCCTCCGGCGTCGCCCAG  
 CAGCGCGTCATACGCCGCGCGTGGGGACGCGCCGGAGTATCGGGCGGCGACGTGCGGAGTC  
 GTCGAGGCCACGGCACCGGCACCCGCCTCGGGGATCCCGTCGAGCTGGGCGCCCTCCTG  
 GGCACGTACGGCGTCGGCCGCGGCGGCGTGGCCCGGTGTCGTGTCGGCAGCGTCAAGGCC  
 10 AACGTCGGCCACGTCCAGGCCGCGGCGGCGTGGTCTGCCGCGGCGGCGCTCAGCGGCCTCGTCGAC  
 CTCGGCCGCGGGCTGGTGGCCCGATGGTCTGCCGCGGCGGCGCTCAGCGGCCTCGTCGAC  
 TGGTCGTCCGGCGGCCTGGTGGTGGCGGACGGGGTCCGCGGCTGGCCGGTGGCGGTGCGAC  
 GCGTCCGCGGGGCGGCGTCTCGGCGTTCGGCGTCAGCGGGACGAAT (SEQ ID NO: 25)  
  
 15 > 6978-7337 Sequence in Module 2  
 GGTGGAGTGTGATGCGGTGGTGTGTCGGTGGTGGGGTTCGGTGTGGGGGTGTTGGA  
 GGGTCGGTCGGGTGCGCCGTCGTTGGATCGGGTGGATGTGGTGCAGCCGGTGTGTTTCGT  
 GGTGATGGTGTGTCGTTGGCGCGGTTGTGGCGGTGGTGTGGGGTGTGTCCTGCGGCGGTGGT  
 GGGTCATTCGCAGGGGGAGATCGCGGCGGCGGTGGTGGCGGGGGTGTGTCGGTGGGTGA  
 20 TGGTGGCGGGTGGTGGCGTTGCGGGCGCGGGCGTTGCGGGCGTTGGCCGG (SEQ ID NO:  
 26)  
 > 27697-27987 Sequence in Module 6  
 GGTGGAGTGTGATGCGGTGGTGTGTCGGTGGTGGGGTTCGGTGTGGGGGTGTTGGA  
 GGGTCGGTCGGGTGCGCCGTCGTTGGATCGGGTGGATGTGGTGCAGCCGGTGTGTTTCGT  
 25 GGTGATGGTGTGTCGTTGGCGCGGTTGTGGCGGTGGTGTGGGGTGTGTCCTGCGGCGGTGGT  
 GGGTCATTCGCAGGGGGAGATCGCGGCGGCGGTGGTGGCGGGGGTGTGTCGGTGGGTGA  
 TGGTGGCGGGTGGTGGCGTTGCGGGCGCGGGCGTTGCGGGCGTTGGCCGG (SEQ ID NO:  
 27)  
 > 27697-27987 Sequence with Codon Changes  
 30 CGTGGAGTGCATGCGGTGTCGAGCGTCGTGCGGCTTCAGCGTGCTGGGCGTCCTGGA  
 GGGCCGCGAGCGGCGCCCGAGCCTGGACCGCGTCGACGTGGTCCAGCCGGTCTGTTTCGT  
 GGTGATGGTGCAGCCTGGCCCGCCTGTGGCGCTGGTGGCGCGTGGTCCCAGCGCCCGTGGT  
 CGGCCACAGCCAGGGCGAGATCGCCGCGCGGTGTCGCGGCGGTCTGAGCGTCGGCGA  
 CGGCGCCCGCGTCGTGGCCCTGCGCGCCCGCGCCCTGCGCGCCCTGGCCGG (SEQ ID NO:  
 35 28)

Three pieces of DNA from the two regions above were synthesized and verified by  
 Retrogen, and the synthesized DNAs were cloned into pCR-Blunt II –TOPO, as  
 shown in the Table 3 below.

40

Table 3. Plasmids containing synthesized DNA

Plasmids	Cloning sites and positions in meg PKS
pKOS97-1613	PstI-BamHI, 26,739 <sup>th</sup> -26,947 <sup>th</sup> base
PKOS97-1622	BamHI-BsmI, 26,947 <sup>th</sup> -27,267 <sup>th</sup> base
PKOS97-1628	SfaNI-FseI, 27,697 <sup>th</sup> - 27,987 <sup>th</sup> base

#### *Assembly of the expression plasmid*

First, ligation of the PstI-BamHI fragment of pKOS97-1613, the BamHI-  
 45 BsmI fragment of pKOS97-1622 and BsmI-PstI linearized pKOS97-90 produced

pKOS97-151. Then, the insertion of the SfaNI-FseI fragment of pKOS97-1628 into pKOS97-151 gave rise to pKOS97-152. Then, the PstI-BlnI fragment of pKOS97-125 was used to replace the PstI-BlnI fragment of pKOS97-90a and produced pKOS97-160.

5 The final expression plasmid (in pRM5) pKOS97-162 was the result of BglII-NheI fragment of pKOS97-160 inserted into BglII-NheI sites of pKOS108-04.

Another expression plasmid pKOS97-152a was made by a four-fragment ligation. The four fragments were a BlnI-XbaI fragment (containing a cos site) of  
10 pKOS97-92a, a BglII-PstI fragment of pKOS97-81, a PstI-BlnI fragment of pKOS97-152, and a BglII-XbaI fragment of pKOS108-04 (as the vector).

Tests of the constructed plasmids showed that the plasmids containing the modified coding sequences were more stable than plasmids containing unmodified coding sequence.

15

### Example 3

#### Construction of Ole-Meg Hybrid PKS

*Construction of pRM1-based pKOS098-48 for the expression of OlePKS modules 1-4.*

20 The 240-bp fragment containing the 3'-end portion of *oleAII* gene (at nt 11210-11452; the first base of the start codon of *oleAII* is nt 1) was PCR amplified with primers N98-38-1 (5'-GAACAACCTCCTGTCTGCGGCCGCG-3') (SEQ ID NO: 29) and N98-38-3 (5'-  
CGGAATTCTCTAGAGTCACGTCTCCAACCGCTTGTCGAGG-3') (SEQ ID  
25 NO: 30). The fragment contains a naturally occurring NotI site at its 5'-end and the engineered XbaI (bold) and EcoRI sites (underline) at its 3'-end following the *oleAII* stop codon. pKOS38-189 was digested with EcoRI and NotI to give five fragments of 8 kb, 5 kb, 4 kb, 2.5 kb and 2 kb. The 8-kb EcoRI-NotI fragment containing *oleAII* gene nt 2961 to nt 11210 and the 240-bp NotI, EcoRI treated  
30 PCR fragment were ligated into litmus 28 at the EcoRI site via a three-fragment ligation to give pKOS98-46. The 8.2-kb EcoRI fragment from pKOS98-46 was cloned into pKOS38-174, a pRM1 derived plasmid containing *oleAI* and nt 1 to nt 2960 of *oleAII* to give pKOS98-48.

*Construction of pSET152-based pKOS98-60 for the expression of megPKS modules 5-6.*

The 360-bp fragment containing nt 1 to nt 366 of *megAIII* was PCR  
5 amplified with primers N98-40-3 (5'-  
TCTAGACTTAATTAAGGAGGACACATATGAGCGA-GAGCAGC-  
GGCATGACCG-3') (SEQ ID NO: 31) and N98-40-2 (5'- AACGCCTCCCAG-  
GAGATCTCCAGCA-3') (SEQ ID NO: 32). A *PacI* site and a *NdeI* site as well  
as the ribosome binding site were introduced at the 5'-end of the *megAI* start  
10 codon. The 360-bp *PacI*-*BglII* fragment was inserted into pKOS108-06 replacing  
the 22-kb *PacI*-*BglII* fragment to yield pKOS98-55. The 10-kb *PacI*-*XbaI*  
fragment containing *megAIII* gene and the annealed oligos N98-23-1 (5'-  
AATTCATAGCCTAGGT-3') (SEQ ID NO: 33) and N98-23-2 (5'-  
CTAGACCTAGGCTATG-3') (SEQ ID NO: 34) were ligated to *PacI* and *EcoRI*  
15 treated pSET152 derivative pKOS98-14 via a three-fragment ligation to give  
pKOS98-60.

Example 4

Conversion of Erythronolides to Erythromycins

20 A sample of a polyketide (~50 to 100 mg) is dissolved in 0.6 mL of  
ethanol and diluted to 3 mL with sterile water. This solution is used to overlay a  
three day old culture of *Saccharopolyspora erythraea* WHM34 (an *eryA* mutant)  
grown on a 100 mm R2YE agar plate at 30°C. After drying, the plate is incubated  
at 30°C for four days. The agar is chopped and then extracted three times with 100  
25 mL portions of 1% triethylamine in ethyl acetate. The extracts are combined and  
evaporated. The crude product is purified by preparative HPLC (C-18 reversed  
phase, water-acetonitrile gradient containing 1% acetic acid). Fractions are  
analyzed by mass spectrometry, and those containing pure compound are pooled,  
neutralized with triethylamine, and evaporated to a syrup. The syrup is dissolved  
30 in water and extracted three times with equal volumes of ethyl acetate. The  
organic extracts are combined, washed once with saturated aqueous NaHCO<sub>3</sub>,  
dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and evaporated to yield ~0.15 mg of product. The  
product is a glycosylated and hydroxylated compound corresponding to

erythromycin A, B, C, and D but differing therefrom as the compound provided differed from 6-dEB.

#### Example 5

5

#### Measurement of Antibacterial Activity

Antibacterial activity is determined using either disk diffusion assays with *Bacillus cereus* as the test organism or by measurement of minimum inhibitory concentrations (MIC) in liquid culture against sensitive and resistant strains of *Staphylococcus pneumoniae*.

10

#### Example 6

#### Evaluation of Antiparasitic Activity

Compounds can initially be screened *in vitro* using cultures of *P. falciparum* FCR-3 and K1 strains, then *in vivo* using mice infected with *P. berghei*. Mammalian cell toxicity can be determined in FM3A or KB cells. Compounds can also be screened for activity against *P. berhei*. Compounds are also tested in animal studies and clinical trials to test the antiparasitic activity broadly (antimalarial, trypanosomiasis and Leishmaniasis).

20

The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments and that the foregoing description and examples are for purposes of illustration and not limitation of the following claims.

25

Claims

1. An isolated nucleic acid comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme.

5

2. The isolated nucleic acid of claim 1, which encodes a PKS open reading frame (ORF) selected from the group consisting of megAI, megAII and megAIII.

10

3. The isolated nucleic acid of claim 1, wherein the PKS domain is selected from the group consisting of a TE domain, a KS domain, an AT domain, an ACP domain, a KR domain, a DH domain, and an ER domain.

15

4. The isolated nucleic acid of claim 1, wherein the nucleic acid comprises the coding sequence for a loading module, a thioesterase domain, and all six extender modules of megalomicin PKS.

20

5. The isolated nucleic acid of claim 1, which encodes a megalomicin modification enzyme that is involved in the conversion of 6-dEB into a megalomicin.

25

6. The isolated nucleic acid of claim 5, which encodes a megalomicin modification enzyme that is involved in the biosynthesis of mycarose, megosamine or desosamine.

30

7. The isolated nucleic acid of claim 1, wherein the nucleic acid codons of homologous regions within the PKS or the megalomicin modification enzyme coding sequence have been changed to reduce or abolish the homology without changing the amino acid sequences encoded by said changed nucleic acid codons.

8. The isolated nucleic acid of claim 1, which isolated nucleic acid fragment hybridizes to a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1.

5 9. A polypeptide, which is encoded by the isolated nucleic acid fragment of claim 1.

10. A recombinant DNA expression vector, comprising the isolated nucleic acid of claim 1 operably linked to a promoter.

10

11. A recombinant host cell, comprising the recombinant DNA expression vector of claim 10.

12. The recombinant host cell of claim 11, which is a *Streptomyces* or  
15 *Saccharopolyspora* host cell.

13. A recombinant host cell of claim 11, which comprises:

a) at least two separate autonomously replicating recombinant DNA expression vectors, each of said vectors comprises a recombinant DNA compound  
20 encoding a megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter; or

b) at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant DNA compound encoding a  
25 megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter.

14. A hybrid PKS that comprises a polypeptide of claim 9 and is composed of at least a portion of a megalomicin PKS and at least a portion of a  
30 second PKS for a polyketide other than megalomicin.



15. The hybrid PKS of claim 14, wherein the second PKS is selected from the group consisting of a narbonolide PKS, an oleandolide PKS, and a DEBS PKS.

5 16. The hybrid PKS of claim 15 that is composed of the megAI and megAII gene products and the oleAIII gene product.

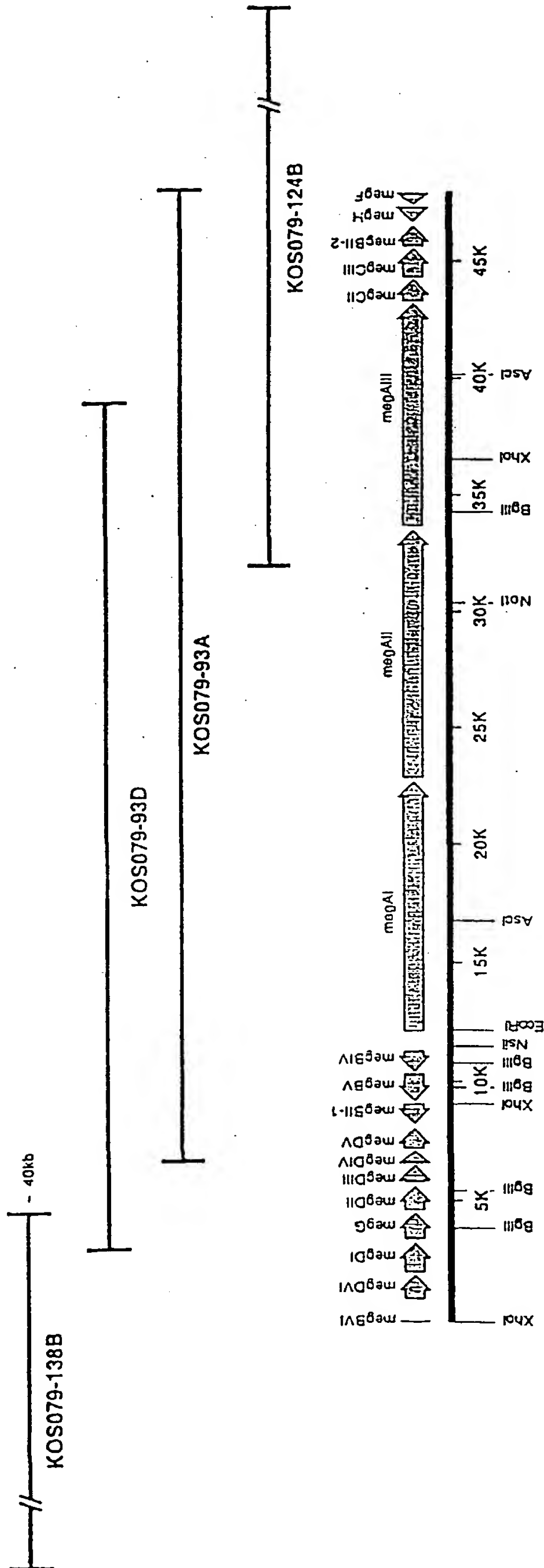
17. The hybrid PKS of claim 16, wherein the KS domain of module 1 of the megAI gene product has been inactivated by mutation.

10

18. A method of producing a polyketide, which method comprises growing the recombinant host cell of claim 11 under conditions whereby the megalomicin PKS domain encoded by the recombinant expression vector is produced and the polyketide is synthesized by the cell, and recovering the  
15 synthesized polyketide.

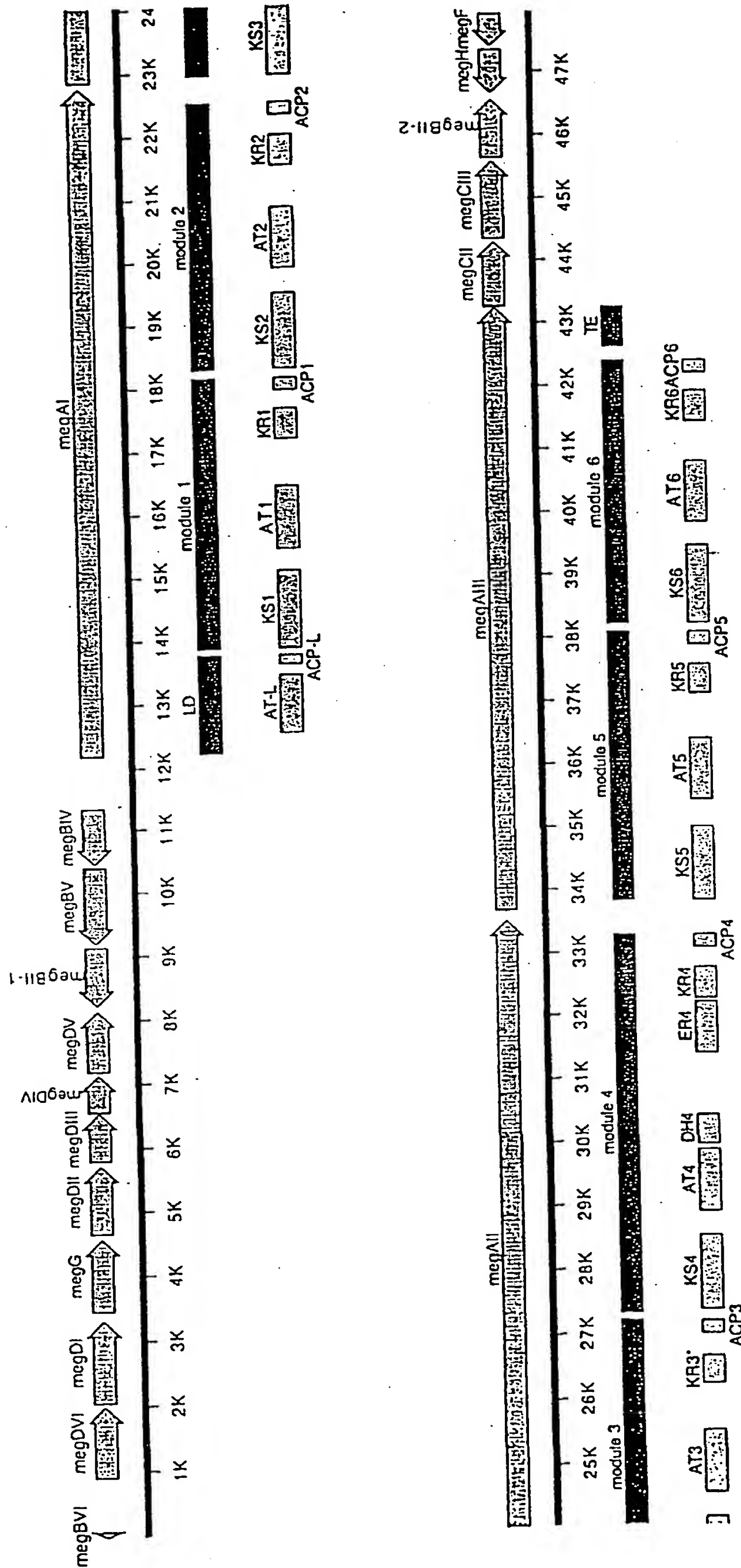
19. A recombinant host cell that comprises a recombinant expression vector that encodes a megalomicin modification enzyme.

20 20. The recombinant host cell of claim 19 that produces megosamine and can attach megosamine to a polyketide, wherein said host cell, in its naturally occurring non-recombinant state cannot produce megosamine.



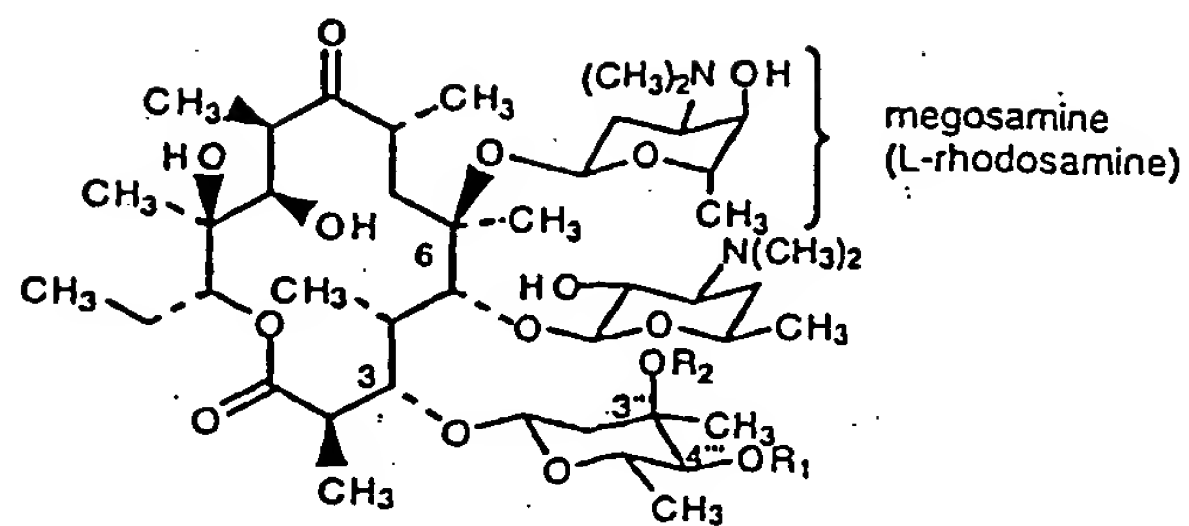
Cosmid Inserts

Figure 1

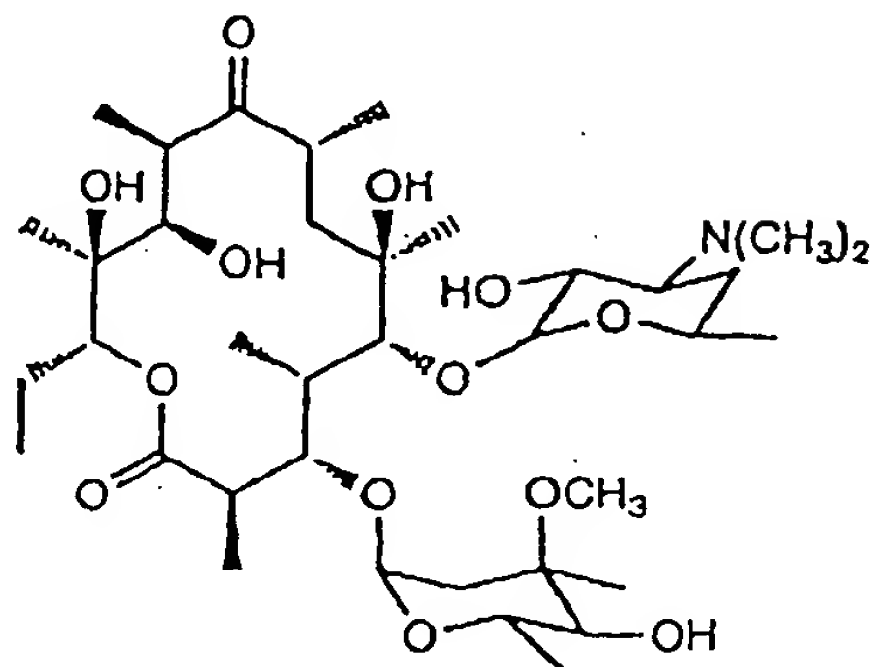
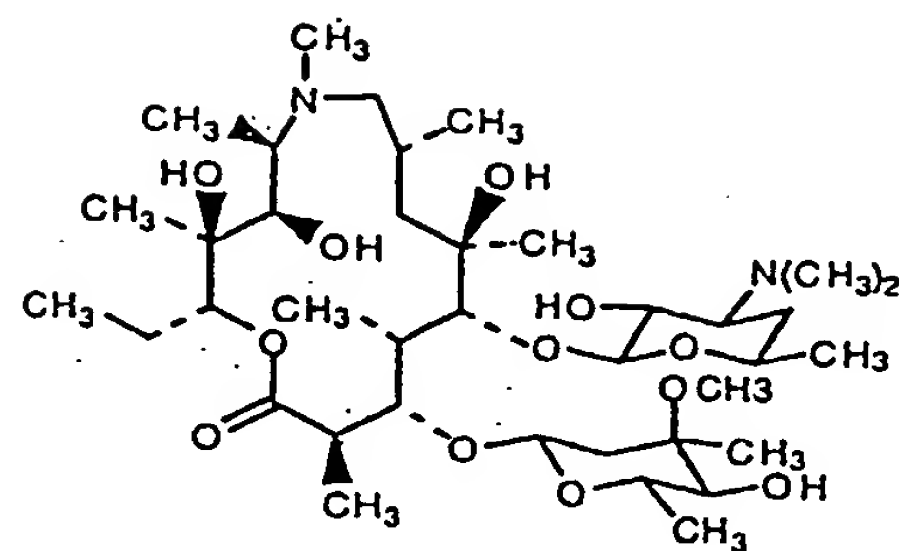


Megalomicin Biosynthetic Genes

Figure 2



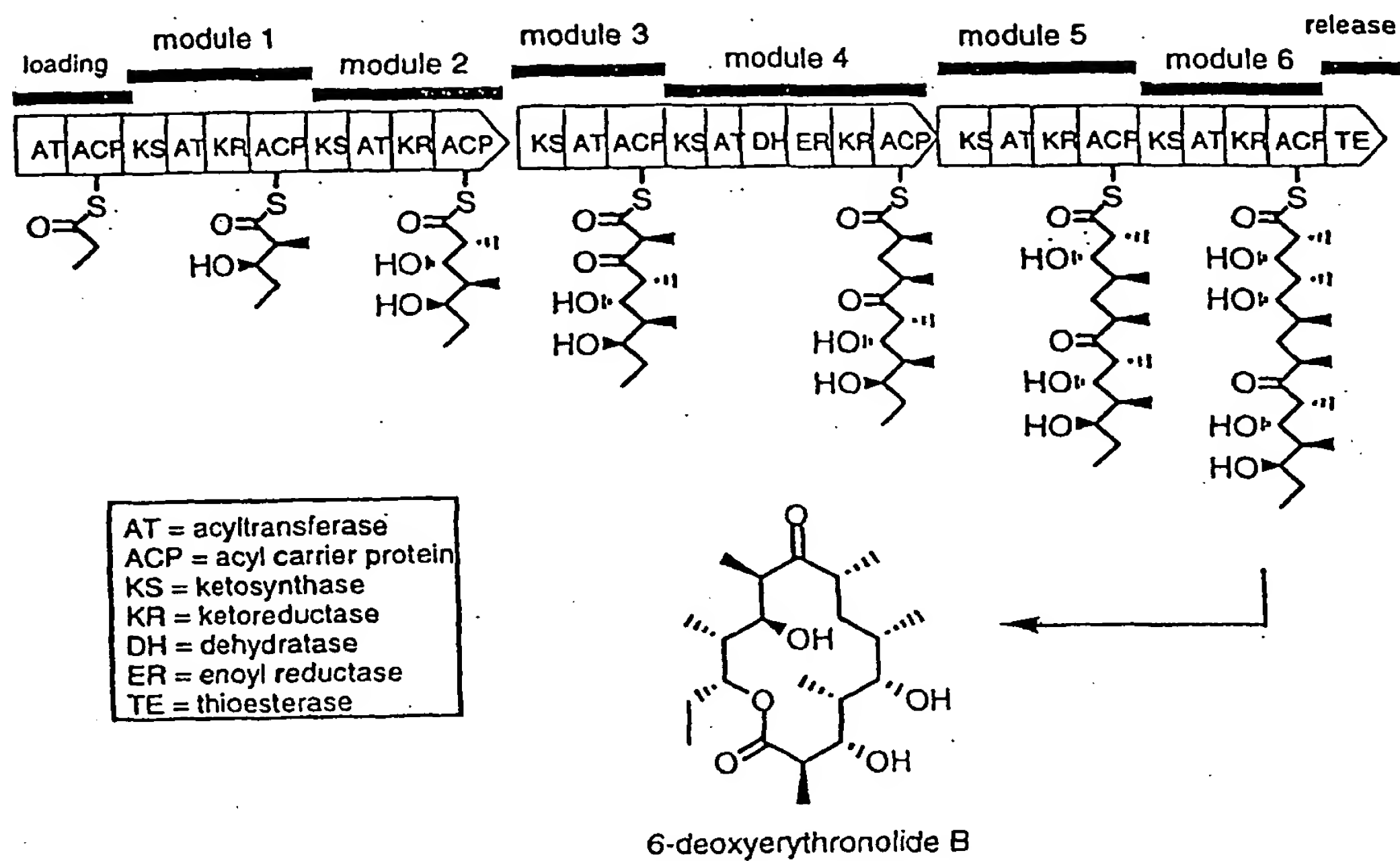
	R <sub>1</sub>	R <sub>2</sub>
Megalomicin A	H	H
B	COCH <sub>3</sub>	H
C1	COCH <sub>3</sub>	COCH <sub>3</sub>
C2	COCH <sub>2</sub> CH <sub>3</sub>	COCH <sub>3</sub>



Erythromycin A

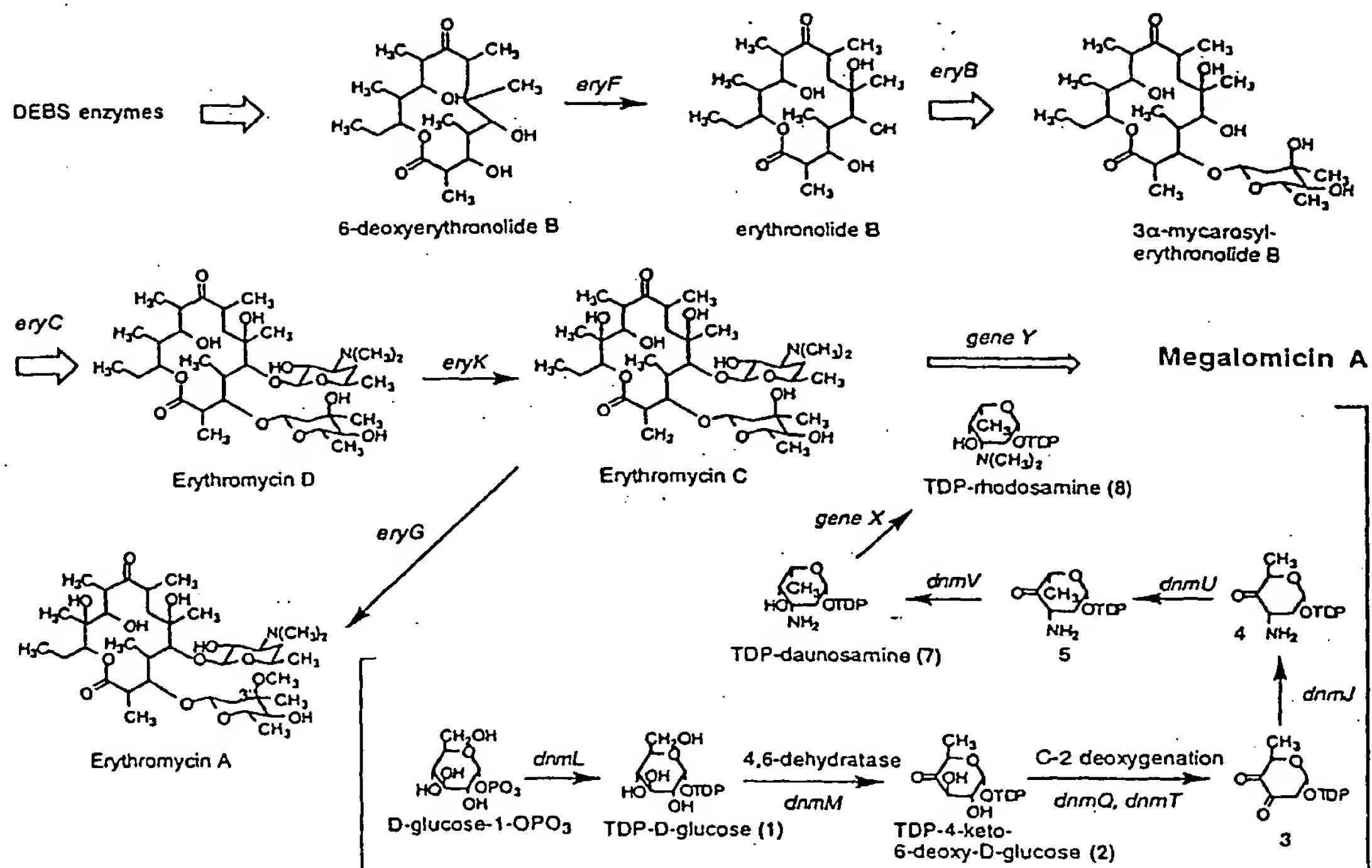
Structures of the Megalomicins and Azithromycin

Figure 3



Biosynthesis of 6-Deoxyerythronolide B (6-dEB), the Aglycone of Erythromycin, by a Modular PKS

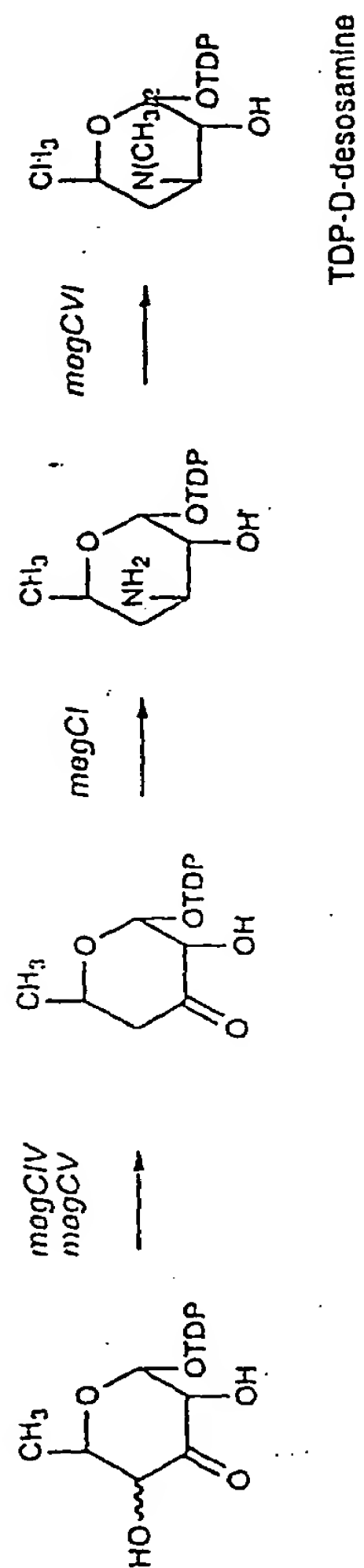
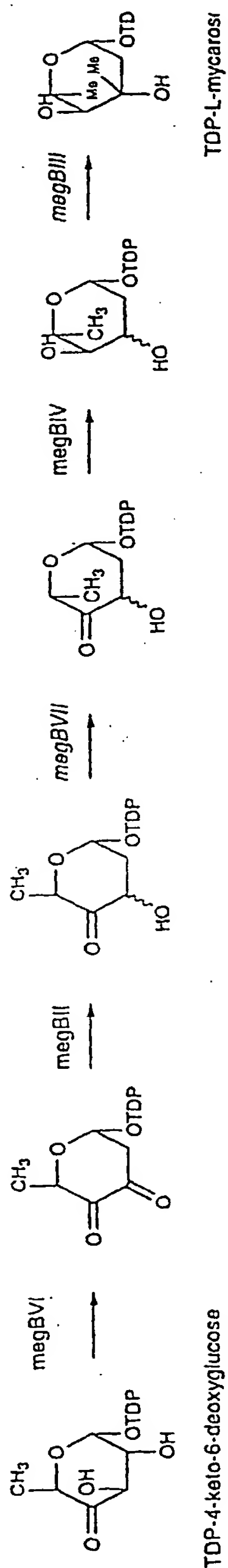
Figure 4



Erythromycin Biosynthetic Pathway and Megalomicin Biosynthesis

Figure 5





## Glycoside Biosynthetic Genes

Figure 6

LOCUS 1 47981 bp DNA 01-MAY-2000  
 DEFINITION Megalomycin biosynthetic gene cluster, polyketide synthase, desosamine, megosamine, and mycarose biosynthesis genes.  
 ACCESSION 1  
 VERSION  
 KEYWORDS  
 SOURCE Micromonospora megalomicea.  
 ORGANISM Micromonospora megalomicea  
 Unclassified.  
 REFERENCE 1 (bases 1 to 47981)  
 AUTHORS Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.  
 TITLE Biosynthesis of the Anti-Parasitic Agent Megalomycin: Transformation of Erythromycin to Megalomycin in Saccharopolyspora erythraea  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 47981)  
 AUTHORS McDaniel, R. and Volchegursky, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center Place, Hayward, CA 94545, USA  
 FEATURES  
 source Location/Qualifiers  
 1..47981  
 /organism="Micromonospora megalomicea"  
 /strain="NRRL3275"  
 /sub\_species="nigra"  
 gene complement(<1..144)  
 /gene="megT"  
 CDS complement(<1..144)  
 /gene="megT"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TDP-4-keto-6-deoxyglucose-2,3-dehydratase"  
 /translation="MGDRVNGHATPESTQSAIRFLTRHGGPPTATDDVHDWLAHRAAEIRLE" (SEQ ID NO: 2)  
 gene 928..2061  
 /gene="megDVI"  
 CDS 928..2061  
 /gene="megDVI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
 /translation="MAVGDRRLGRELQMARGLYWGFANGDLYSMLLSGRDDDPWTWYERLRAAGRGPYASRAGTWVGDHRTAAEVLADPGFTHGPPDAARWMQVAHCPAASWAGPFREFYARTEDAASVTVDADWLQQRCAVLTELGSRFDLVNDFAREVPLALGTAPALKGVDPDRLRSWTSATRVCLDAQVSPQQLAVTEQALTALDEIDAVTGGRDAAVLVGVVAELAANTVGNVAVTELPELAARLADDPETATRVVTEVSRTSPGVHLERRTAASDRRVGGVDVPTGGEVTVVAAANRDPEVFTDPDRFDVDRGGDAEILSSRPGSPRTDLALVATLATAALRAAPVLPRLSRSGPVIRRRRSPVARGLSRCPVEL" (SEQ ID NO: 3)  
 gene 2072..3382  
 /gene="megDI"  
 CDS 2072..3382  
 /gene="megDI"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TDP-megosamine glycosyltransferase"  
 /translation="MRVVFSSMAVNSHLFGLVPLASAFQAAGHEVRVVASPALTDVDTGAGLTAVPVGDDVELVEWHAGQDIVEYMRITLDWVDQSHTTMSWDDLLGMQTTFTPTFFALMSPDSLIDGMVEFCRSWRPDWIVWEPLTFAAPIAARVTGTPHARMLWGPDVATRARQSFLRLLAHQEVEHREDPLAEWFDWTLRRFGDDPHLSFDEELVLGQWTVDPPIPEPLRIDTGVRTVGMRYVPYNGPSVVPWALLREPERRRVCLTLGGSSREHGIGQVSIEMLDIAIDIDAEFVATFDDQQLVGVGSPVANVRTAGFVPMNVLLPTCAATVHHGGTGSWLTAIIHGVPQIILSDADTEVHAKQLQDLGAGLSLPVAGMTAEHLRGAIERVLDEPAYRLGAERM RDGMRTDPSPAQVVGICQDLAADRAARGRQPRRTAEPHLPR" (SEQ ID NO: 4)  
 gene 3462..4634

CDS  
/gene="megY"  
3462..4634  
/gene="megY"  
/codon\_start=1  
/transl\_table=11  
/product="mycarose O-acyltransferase"  
/translation="MVTSTNLDTTARPALNSLTGMRFVAAFLVFFTHVLSRLIPNSYV  
YADGLDAFWQTTGRVGVVSFFFILSGFVLTWSARASDSVWSFWRRRVCKLFPNHLVTAF  
AAVVLFLVTGQAVSGEALIPNLLLIHAWFPALFISFGINPVSWSLACEAFFYLCFPLF  
LFWISGIRPERLWAWAAVFAAIWAVPVVADLLLPSSPPLIPGLEYSATQDWFLYTFF  
ATRSLEFILGIILARILITGRWINVGLLPVLLFPVFFVASLFLPGVYAISSSMILP  
LVLIIASGATADLQQRTEFMRNRVMVWLGDVSFALYMVHFLVIVYGADLLGFSQTEDA  
PLGLALFMIIPFLAVSLVLSWLLYRFVELPVMRNWARPASARRKPATEPEQTPSRR"  
gene 4651..5775 (SEQ ID NO: 5)  
/gene="megDII"  
CDS 4651..5775  
/gene="megDII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"  
/translation="MTTYVWSYLLEYERERADILDAVQKVFASGSLILGQSVENFETE  
YARYHGIAHCVGVNDGTNAVKLALSVGVGRDDEVTVSNTAAPTFLAIDEIGARPVF  
VDVRDEDYLMDDTLVEAAVTPRTKAIVPVHLYGQCVDMTALRELADRRGLKLVEDCAQ  
AHGARRDGRLAGTMSDAAAFSFPYTKVLGAYGDGGAVVTNDDETARALRRLRYGMEE  
VYYVTRTPGHNSRLDEVQAEILRRKLTRLDAYVAGRRAVAQRYVDGLADLQDSHGLEL  
PVVTDGNEHVYVYVVRHPRRDEI IKRLRDGYDISLNISYPWPVHTMTGFAHLGVASG  
SLPVTERLAGEIFSLPMYPSLPHDLQDRVIEAVREVITGL" (SEQ ID NO: 6)  
gene 5822..6595  
/gene="megDIII"  
CDS 5822..6595  
/gene="megDIII"  
/codon\_start=1  
/transl\_table=11  
/product="daunosaminyl-N,N-dimethyltransferase"  
/translation="MPNSHSTTSSTDVAPYERADIYHDFYHGRGKGYRAEADALVEVA  
RKHTPQAATLLDVACGTGSHLVELADSFREVVGVDLSAAMLATAARNDPGRELHQDM  
RDFSLDRRFDVVTCTMFSSGTGYLVDEAELDRAVANLAGHLAPGGTLVVEPWFPETFRP  
GWVGADLVTSGDRRISRMSHTVPAGLPDRITASRMTIHYTVGSPEAGIEHFTVHVMTL  
FARAAEQAFQAGLSCSYVGHDLFSPGLFVGVAEPGR" (SEQ ID NO: 7)  
gene 6592..7197  
/gene="megDIV"  
CDS 6592..7197  
/gene="megDIV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"  
/translation="MRVEELGIEGVFTFTPTQTFADERGTVFGTAYQEDVFVAALGRPLF  
PVAQVSTTRSRRGVVRGVHFTTTPGSMKYVYCARGRAMDFAVDIRPGSPFTGRAEPV  
ELSAESMVGLYLPVGMGHLFVSLEDDTTLVYLMSAGYVPDKERAVHPLDPELALPIPA  
DLDLVMSERDRVAPTLREARDQGILPDYACRAAAHRVVRT" (SEQ ID NO: 8)  
gene 7220..8206  
/gene="megDV"  
CDS 7220..8206  
/gene="megDV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/translation="MVVLGASGFLGSAVTHALADLPVRVRLVARREVVPSPGAVADYE  
THRVDLTPGALAEVVADARAVFPFAAQIRGTSGWRISDDVVAERTNVGLVRDLIAV  
LSRSPHAPVVVFPGSNTQVGRVTAGRVIDGSEQDHPEGVYDRQKHTGEQLLKEATAAG  
AIRATSLRLPPVFGVPAAGTADDRGVVSTMIRRALTGQPLTMWHDGTVRRELLYVTDA  
ARAFVTALDHADALAGRHFLLGTGRSWPLGEVFQAVSRVARHTGEDPVPVVSVPSPA  
HMDPSDLRSVEVDPAFTAVTGWRATVTMAEAVDRTVAALAPRRRAAPSEPS"  
gene complement (8228..9220) (SEQ ID NO: 9)

CDS /gene="megDVII"  
complement(8228..9220)  
/gene="megDVII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 2,3-reductase"  
/translation="MGTTGAGSARVRVGRSALHTSRLWLGTVNFSGRVTDDDALRLMD  
HALERGVNCIDTADIYGWRLYKGHTTELVGRWFAQGGGRREETVLATKVGSEMSERVN  
DGGLSARHIVAACENSLRRLGVDHIDIYQTHHIDRAAPWDEVWQAAEHLVGSGKVGIV  
GSSNLAGWHIAAAQESAARRNLLGMISHQCLYNLAVRHPELDVLPAQAYGVGVFAWS  
PLHGGLLSGVLEKLAAGTAVKSAQGRAQVLLPAVRPLVEAYEDYCRRLGADPAEVGLA  
WVLSRPGILGAVIGPRTPEQLDSALRAAELTLGEEELRELEAIFPAPAVDGPVP"  
gene complement(9226..10479) (SEQ ID NO: 10)  
/gene="megBV"  
CDS complement(9226..10479)  
/gene="megBV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-mycarose glycosyltransferase"  
/translation="MRVLLTSFAHRTHFQGLVPLAWALHTAGHDVRVASQPELTDVVV  
GAGLTSVPLGSDHRLFDISPEAAAQVHRYTTDLDFARRGPPELRSWEFLHGIEEATSRF  
VFPVVNDSFVDELVEFAMDWRPDLVLWEPFTFAGAVAACACGAHARLLWGSDLTGY  
FRSRSQDLRGQRPADDRPDPLGGWLTEVAGRFGLDYSED LAVGQWSVDQLPESFRLET  
GLESVHTRTLPYNGSSVVPQWLRTSDGVRVRCFTGGYSALGITSNPQEFLRTLTLAR  
FDGEIVVTRSGLDPASVPDNVRLVDFVPMNILLPGCAAVIHHGGAGSWATALHHGVPO  
ISVAHEWDCVLRGQRTAELGAGVFLRPDEVADTLWQALATVVEDRSHAENAELRQE  
ALAAPTPAEVVPVLEALAHQHRADR" (SEQ ID NO: 11)  
gene complement(10483..11424)  
/gene="megBIV"  
CDS complement(10483..11424)  
/gene="megBIV"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/translation="MTRHVTL LGVSGFVGSALLREFTTHPLRLRAVARTGSRDQPPGS  
AGIEHLRVLDLLEPGRVAQVVADTDVVVHLVAYAAAGGSTWRSAAATVPEAERNAGIMRD  
LVAALRARP GPAPVLLFASTTQAANPAAPSRYAQHKIEAERILRQATEDGVVDGVILR  
LPAIYGHSGPSGQTGRGVVTAMIRRALAGEPITMWHEGSVRRNLLHVEDVATAFTAAL  
HNHEALVGDVWTPSADEARPLGEIFETVAASVARQTGNPAVPVVSVPPEAEANDFR  
SDDFDSTEFRTL TGWHPRVPLAEGIDRTVAALISTKE" (SEQ ID NO: 12)  
gene 12181..22821  
/gene="megAI"  
CDS 12181..22821  
/gene="megAI"  
/note="polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="megalomycin 6-deoxyerythronolide B synthase 1"  
/translation="MVDVPDLLGTRTPHPGPLPFPWPLCGHNEPELRARARQLHAYLE  
GISEDVAVGAALARETRAQDGPHRAVVVASSVTELTAAALAAQGRPHPSVVRGVA  
RPTAPVVVFLPGQGAQWPGMATRLLAESPVFAAAMRACERAFDEVTDWSL TEVLDSPE  
HLRRVEVVQPALFAVQTS LAALWRSFGVRPDAVLGHSIGELAAAEVCGAVDVEAAARA  
AALWSREMVPLVGRGDMAAVALSPAELAAARVERWDDDDVVPAGVNGPRSVLLTGAPPEI  
ARRVAELAAQGVRAQVVNVSMAAHSAQVDAVAEGMRSALTWFAPGDS DVPYYAGLTGG  
RLDTREL GADHWPRSFRLPVRFDEATRAVLELQPGTFIESSPHPVLAASLQOTLDEVG  
SPAAIVPTLQRDQGLRRFLLAVAQAYTGGVTVDWTAAYPGVTPGHLPSAVAVETDEG  
PSTFDWAAPDHVLRARLLEIVGAETAALAGREVDARATFRELGLDSVLAVQLRTRLA  
TATGRDLHIAMLYDHPTPHALTEALLRGPQEEPGRGEETAHPTEAEPDEPVAVVAMAC  
RLPGGVTSPEEFWELLAEGRDAVGGLPTDRGWDLDLSL FHPDPTRSGTAHQ RAGGFLTG  
ATSFDAAFFGLSPREALAVEPQQRITLSEVLERAGIPPTSLRTSRTGVFVGLIPQ  
EYGPRLAEGGEGVEGYLMTGTTTSVASGRVAYTLGLEGPAISVDTACSSSLVAVHLAC  
QSLRRGESTMALAGGVTVMP TPGMLVDFSRMNSLAPDGRSKAFSAAADGFGMAEGAGM  
LLLERLS DARRHGHPVLAVIRGTAVNSDGASNGLSAPNGRAQVRVIRQALAESGLTPH  
TVDVVETHGTGTRLGDP IEARALSDAYGGDREHPLRIGSVKSNIGHTQAAAGVAGLIK

LVLAMQAGVLPRTLHADEPSPEIDWSSGAISLLQEPAAWPAGERPRRAGVSSFGISGT  
 NAHAIIEEAPPTGDDTRPDRMGPPVVPVLSASTGEALRARAARLAGHLREHPDQDLDD  
 VAYSLATGRAALAYRSGFVPADASTALRILDELAAGGSGDAVTGTARAPQRVVVFVPG  
 QGWQWAGMAVDLLDGDVPFASVLRACADALEPYLDFEIVPFLRAEAQRRTPDHTLSTD  
 RVDVVQPVLFVAVMVSLAARWRAYGVEPAAVIGHSSQGEIAAACVAGALSLLDDAARAVL  
 RSRVIATMPGNGAMASIAASVDEVAARIDGRVEIAAVNGPRAVVVSGDRDDLDRLVAS  
 CTVEGVRAKRLPVDYASHSSHVEAVRDALHAELGEFRPLPGFVPFYSTVTGRWVEPAE  
 LDAGYWFRNLRHRVRFADAVRSLADQGYTTTFLEVSAHPVLTTAIEEIGEDRGDDLAV  
 HSLRRGAGGPVDFGSALARAFVAGVAVDWESAYQGAGARRVPLPTYPFQRRERFWLEPN  
 PARRVADSDDVSSLRYRIEWHPTDPGEPGRLDGTWLLATYPGRADDRVEAARQALES  
 GARVEDLVVEPRTGRVDLVRRLDAVGVPAGVLCFVAEPAEHSPLAVTSLSDTLDL  
 TQAVAGSGRECPIWVVTENAVAVGPFERLRDPAHGALWALGRVVALENPAVWGGLVDV  
 PSGSVAELSRHLGTTLSGAGEDQVALRPDGTYARRWCRAGAGGTGRWQPRGTVLVTGG  
 TGGVGRHVARWLRQGTPLVLASRRGPDADGVEELLTELADLGTRATVTACDVTRE  
 QLRALLATVDDEHPLSAVFHVAATLDDGTVETLTGDRIERANRAKVLGARNLHELTRD  
 ADLDAFVLFFSSSTAAGAPGLGGYVPGNAYLDGLAQRRSEGLPATSVAWGTWAGSGM  
 AEGPVADRFRRHGVMEMHPDQAVEGLRVALVQGEVAPIVVDIRWDRFLLAYTAQRPTR  
 LFDTLDEARRAAPGPDAGPGVAALAGLPVGEREKAVLDLVRTHAAAVLGHASAEQVPV  
 DRAFAELGVDSLALSLRNRLTTATGVRLATTTVFDHPDVRTLACHLAAELGGGSGRE  
 RPGGEAPTVAPTDEPIAIVGMACRLPGGVDSPEQLWELIVSGRDTASAAPGDRSWDPA  
 ELMVSDTTGTRTAFGNFMPGAGEFDAAFFGISPREALAMPQQRHALETTWEALENAG  
 IRPESLRGTDGTGVFVGMSSHQGYATGRPKPEDEVDGYLLTGNTASVASGRIAYVLGLE  
 PAITVDTACSSSLVALHVAAGSLRSGDCGLAVAGGVSVMAGPEVFREFSRQALAPDG  
 RCKPFSDEADGFLGEGSAFVVLQRLSVAVREGRVRLGVVVGSAVNQDGASNGLAAPS  
 GVAQQRVIRRAWGRAGVSGGDVGVEAHGTGTRLGDPVELGALLGTYGVRGGVGPVV  
 VGSVKANVGHVQAAAGVVGVIKVVLLGLGRGLVGPVCRGGLSGLVDWSSGGLVADGV  
 RGWPVGVDGVRGGVSAFGVSGTNAHVVAEAPGSVGAERPVEGSSRGLVGVVGGVV  
 PVVLSAKTETALHAQARRLADHLETHPDVPMTDVWTLTQARQRFDRRAVLLAADRTQ  
 AVERLRGLAGGEPGTGVVSGVASGGGVVFVPGQGGQWVGMARGLLSPVVFVESVVEC  
 DAVVSSVVGFSVLGVLEGRSGAPSLDRVDVVQPVLFVVMVSLARLWRWCGVVPAAVVG  
 HSQGEIAAAVAVAGVLSVGDGARVVALRALARALAGHGGMASVRRGRDDVQKLLDSGP  
 WTGKLEIAAVNGPDVVVSGDPRAVTELVEHCDGIGVPARTIPVDYASHSAQVESLRE  
 ELLSVLAGIEGRPATVPFYSTLTGGFVDGTELDADYWYRNLRHPVRFHAAVEALAARD  
 LTTTFVEVSPHPVLSMAVGETLADVESAVTVGTLETRDDEVERFLTSLAEAHVHGVPVD  
 WAAVLGSGTLVDLPTYPFQGRFRLHPDRGPRDDVADWFHRVDWTATATDGSARLDGR  
 WLVVVPEGYTDDGWVVEVRAALAAGGAEPVTTVEEVTDRVGDSDAVVSMLGLADDGA  
 AETLALLRRLDAQASTTPLWVVTVGAVAPAGPVQRPEQATVWGLALVASLERGHRWTG  
 LLDLPQTPDPQLRPLVEALAGAEDQVAVRADAVHARRIVPTPVTGAGPYTAPGGTIL  
 VTGGTAGLGAVTARWLAERGAEHLALVSRRGPGTAGVDEVVRDLTGLGVRVSVHSCDV  
 GDRESVGALVQELTAAGDVVRGVVHAAGLPQQVPLTMDPADLADVAVKVDGAVHLA  
 DLCPEAELFLLFSSGAGVWGSARQAGAYAAGNAFLDAFARHRRDRGLPATSVAWGLWAA  
 GGMTGDQEAVSFLRERGVRPMSVPRALEALERVLTAGETAVVVADVDWAAFAESYTS  
 RPRPLLHRLVTPAAAVGERDEPREQTLRDLRALPRAERSAELVRLVRRDAAAVLGSD  
 AKAVPATTPFKDLGFDLSAAVRFNRRLAAHTGLRLPATLVFEHPNAAVADLLHDLRG  
 EAGEPTPVRVSVGAGLALEQALPDASDTERVELVERLERMLAGLRPEAGAGADAPTAG  
 DDLGEAGVDELDDALERELDAR" (SEQ ID NO: 13)

misc\_feature 12505..13470  
 /gene="megAI"  
 /function="AT-L"  
 misc\_feature 13576..13791  
 /gene="megAI"  
 /function="ACP-L"  
 misc\_feature 13849..15126  
 /gene="megAI"  
 /function="KS1"  
 misc\_feature 15427..16476  
 /gene="megAI"  
 /function="AT1"  
 misc\_feature 17155..17694  
 /gene="megAI"  
 /function="KR1"  
 misc\_feature 17947..18207  
 /gene="megAI"  
 /function="ACP1"



misc\_feature 18268..19548  
/gene="megAI"  
/function="KS2"  
misc\_feature 19876..20910  
/gene="megAI"  
/function="AT2"  
misc\_feature 21517..22053  
/gene="megAI"  
/function="KR2"  
misc\_feature 22318..22575  
/gene="megAI"  
/function="ACP2"  
gene 22867..33555  
/gene="megAII"  
CDS 22867..33555  
/gene="megAII"  
/note="polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="megalomicin 6-deoxyerythronolide B synthase 2"  
/translation="MTDNDKVAEYLRRATLDLRAARKRLRELQSDPIAVVGMACRLPG  
GVHLPQHLWDLRLRQGHETVSTFPTGRGWDLAGLFHPDPDHPGTSYVDRGGFLDDVAGF  
DAEFFGISPREATAMDPQORLLETSWELVESAGIDPHSLRGTPPTGVFLGVARLGYGE  
NGTEAGDAEGYSVTGVAPAVASGRISYALGLEGPSISVDTACSSSLVALHLAVESLRL  
GESSLAVVGGAAMATPGVFVDFSRQRALAADGRSKAFGAAADGFGFSEGVSLVLLER  
LSEAESNGHEVLAVIRGSALNQDGASNGLAAPNGTAQRKVIRQALRNCGLTPADVDAV  
EAHGTGTTLGDPIEANALLDTYGRDRDPDHLWLGSVKSNIIGHTQAAAGVTGLLKMVL  
ALRHEELPATLHVDEPTPHVDWSSGAVRLATRGPRWRGDRPRRAGVSAFGISGTNAH  
VIVEEAPERTTERTVGGDVGPVPLVVSARSAAALRAQAAQVAELVEGSDVGLAEVGRS  
LAVTRARHEHRAAVVASTRAEAVRGLREVAAVEPRGEDTVTGVAETSGRTVVFLFPGQ  
GSQWVGMAELLD SAPAFADTIRACDEAMAPLQDWSVSDVLRQEPGAPGLDRVDVQP  
VLFVAVMSLARLWQSYGVTPAAVVGHSQGEIAAAHVAGALS LADAARLVVGRSRLRS  
LSGGGMSAVALGEAEVRRRLRSWEDRISVAAVNGPRSVVVAGEPEALREWGREREAE  
GVRVREIDVDYASHSPQIDRVDELTVTGEIEPRSAEITFYSTVDVRAVDGTDLDAG  
YWYRNLRETVRFADAMTRLADSGYDAFVEVSPHPVVVSAVAEAVEEAGVEDAVVVGTL  
SRGDGGPGAFLRSAATAHCAGVDVDWTPALPGAATIPLPYTPFQRPYWLRSAPAPA  
SHDLAYRVSWTPITPPGDGVLGDWLVVHPGGSTGWVDGLAAAITAGGGRVVAHPVDS  
VTSRTGLAEALARRDGTFRGVLSWVATDERHVEAGAVALLTLAQUALGDAGIDAPLWCL  
TQEAVRTPVDGDLARPAQAALHGFAQVARLELARRFGGVLDLPATVDAAGTRLVA AVL  
AGGGEDVAVRGDRLYGRRLVRATLPPPGGGFTPHGTVLVTGAAGPVGGRLARWLAER  
GATRLVLPGAHPGEELLTAIRAAGATAVCEPEAEALRTAIGGELPTALVHAETLTNF  
AGVADADPEDFAATVAAKTALPTVLAEVLGDHRLEREVYCSSVAGVWGGVGMAYAAAG  
SAYLDALVEHRRARGHASASVAWTPWALPGA VDDGRLRERGLRSLDVADALGTWERLL  
RAGAVSVAVADVDWSVFTEGFAAIRPTPLFDELDRRGDPDGAPVDRPGEPAGEWGRR  
IAALSPQEQRETLTLVGETVAEVLGHETGTEINTRAFSELGLDSLGSMA LRQRLAA  
RTGLRMPASLVFDHPTVTALARYLRLRVGDS DPTPVRVFGPTDEAE PVAVVGIGCRF  
PGGIATPEDLWRV VSEGTSITTGFPTDRGWDLRRLYHPDPDHPGTSYVDRGGFLDGAP  
DFDPGFFGITPREALAMPQORLTLEIAWEAVERAGIDPETLLGSDTG VFGVMNGQSY  
LQLLTGEGDRLNGYQGLGNSASVLSGRVAYTFGWEGPALTVDTACSSSLVAIHLAMQS  
LRRGECSLALAGGVTVMADPYTFVDFSAQRGLAADGRCKAFSAQADGFALAEGVAAALV  
LEPLSKARRNGHQVLAVLRGS AVNQDGASNGLAAPNGPSQERVIRQALTASGLRPADV  
DMVEAHGTGTELGDPIEAGALIAAYGRDRDRPLWLGSVKTNIIGHTQAAAGAAGVIKAV  
LAMRHGVLPRSLHADELS PHIDWADGKVEVLREARQWPPGERPRRAGVSSFGVSGTNA  
HVIVEEAPAEPDPEPVPAAPGGPLPFVLHGRSVQTVRSQARTLA EHLRTTGHRDLADT  
ARTLATGRARFDVRAAVLGTDRGVCAALDALAQDRPSPDVVAPAVFAARTPVLVFP  
QGSQWVGMA RDLLDSSEVFAESMGRCAEALS PYTDWDLDDVVRGVGD PDYPDRVDVLQ  
PVLFAVMVSLARLWQSYGVTPGAVVGHSQGEIAAAHVAGALS LADAARVVALRSRVL  
ELDDQGMVSVGTSRAELDSVLRWDGRVAVAAVNGPGTLVVAGPTAELDEFLAVAEA  
REMRPRRIAVRYASHSPEVARVEQRLAAELGTVTAVGGTVPLYSTATGDLLDTTAMDA  
GYWYRNL RQPVLF EHAVRSLLERGFETFIEVSPHPVLLMAVEETAEDAERPVTGVPTL  
RRDHDGPSEFLRNLLGAHVHGV DDLRPAVAHGRVLVDLPTYPFDRQRLWPKPHRRADT  
SSLGVRDSTHPLLHAAVDVPGHGGAVFTGRLSPDEQQWLTQHVVGGRNLPVGSVLVDL  
ALTAGADVGVPLVLEELVLQQLPLVLTAA GALLRLSVGA ADEDGRRPVEIHAAEDVSDPA  
EARWSAYATGTLAVGVAGGGRDGTQWPPPGATALTLTDHYDTLAE LGYEYGP AFQALR



AAWQHGDVVYAEVSLDAVEEGYAFDPVLLDAVAQTFGLTSRAPGKLPFAWRGVTLHAT  
GATAVRVVATPAGPDAVALRVTDPTGQLVATVDALVVRDAGADRDQPRGRDGDHLRLE  
WVRLATPDPTPAAVVHVAADGLDDLLRAGGPAPQAVVVRYRPGDDPTAEARHGVLWA  
ATLVRRWLDDDRWPATTLVVATSAGVEVSPGDDVPRPGAAAVWGVLRCAQAESPDRFV  
LVDGDPETPPAVPDNPQLAVRDGAVFVPRLTPLAGPVPAVADRAYRLVPGNGGSIKAV  
AFAPVPDADRPLAPEEVRVAVRATGVNFRDVLLALGMYPEPAEMGTEASGVVTEVGS  
VRRFTPGQAVTGLFQGAFGPVAVADHRLTTPVPDGVRAVDAAAVPIAFTTAHYALHDL  
AGLQAGQSVLVHAAAGGVGMAAVALARRAGAEVFATASPAKHPTLRALGLDDDDHIASS  
RESGFGERFAARTGGRGVDVVLNSLTGDLDESARLLADGGVFVEMGKTDLRPAEQFR  
GRYVPFDLAEAGPDRLGEILEEVVGLLAAGALDRLPVSVWELSAAPAALTHMSRGRHV  
GKLVLTPQAPVHPDGTVLVTGGTGTGLRLVARHLVTGHGVPHLLVASRRGPAAPGAAE  
LRADVEGLGATIEIVACDTADREALAALLDSIPADRPLTGVVHTAGVLADGLVTSIDG  
TATDQVLRAKVDAAWHLHDLTRDADLSFFVLFSASAASVLAGPGQGVYAAANGVLNALA  
GQRRALGLPAKALGWGLWAQASEMTSGLGDRIARTGVAALPTERALALFDAALRSAGE  
VLFPLSVDRSALRAEYVPEVLRGAVRSTPRAANRAETPGRGLLDRLVGAPETDQVAA  
LAELVRSHAAAVAGYDSADQLPERKAFKDLGFDLSAAVELRNRLGVTTGVRLPSTLVF  
DHPTPLAVAHLRSELFADSAPDVGVGARLDDLERALDALPDAQGHADVGARLEALLR  
RWQSRPPPETEPVTISDDASDDELFSMLDRRLGGGGDV" (SEQ ID NO: 14)

misc\_feature 22957..24237  
/gene="megAII"  
/function="KS3"

misc\_feature 24544..25581  
/gene="megAII"  
/function="AT3"

misc\_feature 26230..26733  
/gene="megAII"  
/function="KR3 (inactive)"

misc\_feature 26998..27258  
/gene="megAII"  
/function="ACP3"

misc\_feature 27393..28590  
/gene="megAII"  
/function="KS4"

misc\_feature 28897..29931  
/gene="megAII"  
/function="AT4"

misc\_feature 29953..30477  
/gene="megAII"  
/function="DH4"

misc\_feature 31396..32244  
/gene="megAII"  
/function="ER4"

misc\_feature 32257..32799  
/gene="megAII"  
/function="KR4"

misc\_feature 33052..33312  
/gene="megAII"  
/function="ACP4"

gene 33666..43271  
/gene="megAIII"

CDS 33666..43271  
/gene="megAIII"  
/note="polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="megalomycin 6-deoxyerythronolide B synthase 3"  
/translation="MSESSGMTEDRLRRYLKRTVAELDSVTGRLDEVEYRAREPIAVV  
GMACRFPGGVDSPEAFWEFIRDGGDAIAEAPTDRGWPPAPRRLGGLLAEPGAFDAF  
FGISPREALATDPQQLMLEISWEALERAGFDPSLRGSAGGVFTGVGAVDYGPRPDE  
APEEVLGYVGIGTASSVASGRVAYTLGLEGPAVTVDACSSGLTAVHLAMESLRDEC  
TLVLAGGVTVMSSPGAFTEFRSQGLAEDGRCKPFSRAADGFLAEGAGVLVLQRLSV  
ARAEGRPVLAVLRGSAINQDGASNGLTAPSGPAQRRVIRQALERARLRPVDVDYVEAH  
GTGTRLGDPIEAHALLDITYGADREPGRPLWVGSVKSNIGHTQAAAGVAGVMKTVLALR  
HREIPATLHFDEPSPHVDWRGAVSVVSETRPWPVGERPRRAGVSSFGISGTNAHVIV

EEAPSPQAADLDPTPGPATGATPGTDAAPTAEPGAEEAVALVFSARDERALRAQAARLA  
 DRLTDDPAPSLRDTAFTLVTRRATWEHRAVVVGGGEEVLAVAGGRPVDPVAVSGR  
 ARAGRRLVLFPGQGAQWQGMARDLLRQSPPTFAESIDACERALAPHVDWSLREVLDGE  
 QSLDPVDVVPVLFVAVVSLARLWQSYGVTPGAVVGHSSQGEIAAAHVAGALSLADAAR  
 VVALRSRVLRRLLGGHGGMASFGLHPDQAAERIRFAGALTVASVNGPRSVVLAGENGP  
 LDELIAECEAEGVTARRIPVDYASHSPQVESLREELLAALAGVRPVSAGIPLYSTLTG  
 QVIETATMDADYWFANLREPVRFDATRQLAEAGFADFVEVSPHPVLTVGVEATLEAV  
 LPPDADPCVTGTLRRERGGLAQFHTALAEAYTRGVEVDWRTAVGEGRPVDLPVYPFQR  
 QNFWLPVPLGRVPDGTGDEWRYQLAWHPVDLGRSSLAGRVLVVTGAAPPAWTDVVRDG  
 LEORGATVVLCTAQSRRRIGAALDAVDGTALSTVVSLLALAEGGAVDDPSLDTLALVQ  
 ALGAAGIDVPLWLVTDRDAAVTVGDDVDPAQAMVGGGLGRVVGVESPARWGGLVDLREA  
 DADSARSLAAILADPRGEEQFAIRPDGVTVARLVPAAPARAAGTRWTPRGTVLVTGGTG  
 GIGAHLARWLAGAGAEHLVLLNRRGAEEAAGAADLDELVALGTGVTITACDVADRDL  
 AAVLDAARAQGRVVTAVFHAAGISRSTAVQELTESEFTEITDAKVRGTANLAELCP  
 DALVLFSSNAAVWGSPLASYAAGNAFLDAFARRGRRSGLPVTSTIAWGLWAGQNMAGT  
 EGGDYLRSQLRAMDPQRAIEELRTTLDAGDPWVSVDLDRERFVELFTAARRRPLFD  
 ELGGVRAGAEETGQESDLARRLASMPAEERHEHVARLVRAEVAAVLGHGTPTVIERDV  
 AFRDLGFDSMTAVDLNRRLAAVTGVRVATTIVFDHPTVDRLTAHYLERLVGEPEATTP  
 AAVVVPQAPGEADEPIAIVGMACRLAGGVRTPDQLWDFIVADGDAVTEMPSDRSWDL  
 ALFDPDPERHGTSYSRHGAFLDGAADFDAFFGISPREALAMPQQRQVLETTWELFE  
 NAGIDPHSLRGTDGTGVFLGAAVQGYQNAQVPKESEGYLLTGSSAVASGRIAYVLGL  
 EGPAITVDTACSSSLVALHVAAGSLRSGDCGLAVAGGVSVSMAGPEVFTEFSRQALAP  
 DGRCKPFSQADGFGFAEGVAVVLLQRLSVAVREGRRVLGVVVGSAVNQDGASNGLAA  
 PSGVAQQRVIRRAWGRAGVSGGDVGVEAHGTGTRLGDPVELGALLGTYGVRGGVGP  
 VVVGSVKANVGHVQAAAGVVGVIKVVLLGLGRGLVGPVCRGGLSGLVDWSSGGLVAD  
 GVRGWPGVDGVRGGVSAFGVSGTNAHVVAEAPGSVVGAEERPVEGSSRGLVGVAGG  
 VVPVLSAKTETALTELARRLHDAVDDTVALPAVAATLATGRAHLPYRAALLARDHDE  
 LRDLRAFTTGSAAPGVVSGVASGGGVVVFVPGQGGQWGMARGLLSVPVFVESVVEC  
 DAVVSSVVGFSVLGVLEGRSGAPSLDRVDVVPVLFVVMVSLARLWRWCGVVPAAVVG  
 HSQGEIAAAVAVGVLSVGDGARVVALRARALRALAGHGMVSLAVSAERARELIAPWS  
 DRISVAAVNSPTSVVVSGDPQALALVAHCAETGERAKTLPVDYASHSAHVEQIRDTI  
 LTDLADVTARRPDVALYSTLHGARGAGTMDARYWYDNLRSVRFDEAVEAAVADGYR  
 VVEMSPHPVLTAAVQEIIDETVAIGSLHRDTGERHLVAELARAHVHGVPVDWRILP  
 ATHPVPLPNYPFEATRYWLAPTAADQVADHRYRVDWRPLATTPAELSGSYLVFGDAPE  
 TLGHSVEKAGGLLVPAAPDRESLAVALDEAAGRLAGVLSFAADTATHLARHRLLEA  
 DVEAPLWLVTSGGVALDDHDPIDCDQAMVWGI GRVMGLETPHRWGGLVDVTVEPTAED  
 GVVFAALLAADDHEDQVALRDGIRHGRRLVRAPLTTRNARWTPAGTALVTGGTGALGG  
 HVARYLARSVTDLVLLSRSGPDAPGAELAAELADLGAEPREACDVTGPRRLRALV  
 QELREQDRPVRIVVHTAGVPDSRPLDRIDELESVSAKVTGARLLDELCPDADTFVLF  
 SSGAGVWGSANLGAYAAANAYLDALAHRRRQAGRAATSVAVGAWAGDGMATGDLGLT  
 RRGLRAMAPDRALRACTRRWTTHTDCVSVADVDWDRFAVGFTAARPRPLIDELVTSAP  
 VAAPTAAAAPVPAMTADQLLQFTRSHVAAIILGHQDPDAVGLDQPFTELGFDSLTAAGL  
 RNQLQATGRTLPAALVFQHTVRRRLADHLAQQLDVGTPAVEATGSVLRDGYRRAGQT  
 GDVRSYLDLLANLSEFRERFTDAASLGGQLELVDLADGSGPVTVICCAGTAALSGPHE  
 FARLASALRGTVPVRALAQPGYEAGEPVPASMEAVLGVAQADAVLAAQGDTPFVLVGH  
 AGALMAYALATELADRGHPPRGVLLDVYPGPHQEAHVHAWLGELTAALFDHETVRMDD  
 TRLTALGAYDRLTGRWRPRDTGLPTLVVAASEPMGEWPDDGWQSTWPFHDRVTVPGD  
 HFSMVQEHADAIAHDAWLSGERA" (SEQ ID NO: 15)

misc\_feature 33780..35027  
 /gene="megAIII"  
 /function="KS5"  
 misc\_feature 35385..36419  
 /gene="megAIII"  
 /function="AT5"  
 misc\_feature 37068..37604  
 /gene="megAIII"  
 /function="KR5"  
 misc\_feature 37860..38120  
 /gene="megAIII"  
 /function="ACP5"  
 misc\_feature 38187..39470  
 /gene="megAIII"  
 /function="KS6"  
 misc\_feature 39795..40811

misc\_feature /gene="megAIII"  
/function="AT6"  
41406..41936  
misc\_feature /gene="megAIII"  
/function="KR6"  
42168..42425  
misc\_feature /gene="megAIII"  
/function="ACP6"  
42585..43271  
gene /gene="megAIII"  
/function="TE"  
43268..44344  
CDS /gene="megCII"  
43268..44344  
/gene="megCII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyglucose 3,4-isomerase"  
/translation="MNTTDRVLGRRLQMI RGLYWGYSNGDPYPMLLCGHDDDPHRW  
YRGLGGSGVRRSRTETWVVTDHATAVRVLDDPTFTRATGRTPEWMRAAGAPASTWAQP  
FRDVHAASWDAELPDPQEVEDRLTGLLPAPGTRLDLVRDLAWPMASRGVGADDPDVL R  
AAWDARVGLDAQLTPQPLAVTEAAIAAVPGDPHRRALFTAVEMTATAFVDAVLAVTAT  
AGAAQRLADDPDVAARLVAEVLRLHPTAHLERRTAGTETVVGEHTVAAGDEVVVVAA  
ANRDAGVFADPDRLLDPDRADADRALSQRGHPGRLEELVVVLTTAALRSVAKALPGLT  
AGGPVVRRRRSPVLRATAHCPVEL" (SEQ ID NO: 16)  
gene 44355..45623  
/gene="megCIII"  
CDS 44355..45623  
/gene="megCIII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-desosamine glycosyltransferase"  
/translation="MRVVFSSMASKSHLFGLVPLAWAFRAAGHEVRVVASPALTDIT  
AAGLTAVPVGTDVDLVDLDFMTHAGYDIIDYVRSDFSERDPATSTWDHLLGMQTVLTPT  
FYALMSPDSLVEGMISFCRSWRPDWSSGPQTFAASIAATVTGVAHARLLWGPDITVRA  
RQKFLGLLPGQPAAHREDPLAEWLTWSVERFGGRVPQDVEELVVGQWTIDPAPVGMRL  
DTGLRTVGMRYVDYNGPSVVPDWLHDEPTRRRVCLTLGISSRENSIGQVSVDLLGAL  
GDVDAEIIATVDEQQLEGVAHV PANIRTVGFVPMHALLPTCAATVHHGGPGSWHTAAI  
HGVPQVILPDGWD TGVRARTEDQGAGIALPVPELTSDQLREAVRRVLDDPAFTAGAA  
RMRADMLAEPSPAEVVDVCAGLVGERTAVG" (SEQ ID NO: 17)  
gene 45620..46591  
/gene="megBII"  
CDS 45620..46591  
/gene="megBII"  
/codon\_start=1  
/transl\_table=11  
/product="TDP-4-keto-6-deoxyglucose 2,3 dehydratase"  
/translation="MSTDATHVRLGRCALLTSRLWLGTAAAGQDDADAVRLLDHARS  
RGVNCLDTADDDSA TSAQVAEESVGRWLAGDTGRREETVLSVTVGVPVPPGGQVGGGGL  
SARQIIASCEGSLRRLGVDHVDVLHLPRVDRVEPWQAVDALVAAGKVCYVGSSG  
FPGWHIVAAQEHAVRRHRLGLVSHQCRYDLTSRHPELEVLPAAQAYGLGVFARPTRLG  
GLLG DGPGAAAARASGQPTALRS AVEAYEVFCRDLGEHPAEVALAWVLSRPGVAGAV  
VGARTPGRLD SALRACGVALGATELTALDGIFPGVAAAGAAPEAWLR" (SEQ ID NO: 18)  
gene complement(46660..47403)  
/gene="megH"  
CDS complement(46660..47403)  
/gene="megH"  
/note="putative thioesterase"  
/codon\_start=1  
/transl\_table=11  
/product="TEII"  
/translation="MNTWLRRFGSADGHRARLYCFPHAGAAADSYLDLARALAPEVDV  
WAVQYPGRQDRRDERALGTAGEIADEVA AVLRLDLVGEVPPFALFGHSMGALVAYETARR  
LEARPGVRPLRLFVSGQTAPRVHERRTDLPDEDGLVEQMRR LGVSEAALADQGLLDMS

LPVLRADHRVLRSYAWQAGPPLRAGITTLCGDTDPLTTVEDAQRWLPYSVVPGRTRTF  
 PGGHFYLAHDVGEVAESVAPDLLRLTPTG" (SEQ ID NO: 19)  
 gene complement(47411...>47981)  
 /gene="megF"  
 CDS complement(47411...>47980)  
 /gene="megF"  
 /codon\_start=1  
 /transl\_table=11  
 /product="C-6 hydroxylase"  
 /translation="IRVQDDADRLSRDELTSIALVLLLAGFEASVSLIGIGTYLLLT  
 HPDQLALVRKDPALLPGAVEEILRYQAPPETTTTFATAEVEIGGV TIPAYSTVLIANG  
 AANRDPGQFPDPDRFDVTRDSRGHLTFGHGIHYCMGRPLAKLEGEVALGALFDRFPKL  
 SLGFPSDEVVWRRSLLLRLGIDHLPVRPNG" (SEQ ID NO: 20)

BASE COUNT 5962 a 16875 c 18045 g 7099 t

ORIGIN

```

1  ctcgagccga  tgctcggcgg  cgcggtgggc  caaccagtcg  tggacgtcgt  cggtaggcggt
61  gggaggtccg  ccgtgccgag  tcaggaaacg  tattgccgat  tgtgtggatt  ccggagtcgc
121 atgaccgttg  acccgatccc  ccatacgect  ctcccgtgat  gtcgtgggcg  gtccgtgcgg
181 taccgcccgg  actgacattc  gtcgatcaag  accccgcccc  gtgtagggct  ccgcccgcga
241 cgggagaagg  tccgtcgaac  aacttcgggg  tgaccggctc  ccggcgctcg  tgaaacgggc
301 gtcggagcac  ccgatcattg  ctgtcgggtg  acttcctaac  tgtcggcgcg  cacatctttc
361 tgaccggtgt  gttccgtggg  atgacgcgtt  cccggcccgt  ctggaactgt  gcgtgggact
421 gaccggttgc  ggcgtgtttt  cgcccgtttc  cgaactgcgg  attcgtcgat  cgcgccagggt
481 ggagcgggtg  gctgaccggg  atgatctgca  atcatggcgc  tcaatgacga  tctctttagt
541 catggtccgc  gccgagggtc  cgacaggccc  gaaacgcccc  gcattccagg  tgttcgacga
601 cgtcgacatc  accgtgcaag  ccgcgatgac  accgacacca  cgccatgctg  gtgccgcact
661 ggaagggtgg  cgcgatcagg  gaaatggccg  tgtcactaga  cagacgcaa  acagctgtcc
721 gggcctgcgg  aaacagcatc  gatctgcgtc  agccgttcat  tgccccggcg  gcaccgcctt
781 ggaaatccgt  gccaccggtc  gtccgcagtg  acgatcgccg  acccgggttt  cgagacagca
841 ggtagtaggc  gatgcaggcg  ttctcgtctc  cgccggacgc  gtcgcactag  gtggaatccg
901 tcacagtctt  caatccggga  gcgttctatg  gcagttggcg  atcgaaggcg  gctgggcccg
961 gagttgcaga  tggccccggg  tctctacttg  gggttcgggt  ccaacggcga  tctgtactcg
1021 atgctcctgt  ccggacggga  cgacgacccc  tggacctggt  acgaacgggt  gcgggcccgc
1081 ggacggggac  cgtacgccag  tcggggccga  acgtgggtgg  tcggtgacca  ccggaccgcc
1141 gccgaggtgc  tcgccgatcc  gggttccacc  cacggcccgc  ccgacgctgc  ccggtggatg
1201 caggtggccc  actgcccggc  ggcttctctg  gccggcccct  tccgggagtt  ctacgcccgc
1261 accgaggacg  cggcgtcggg  gacagtggac  gccgactggc  tccagcagcg  gtgcgccagg
1321 ctggtgaccg  agctggggtc  gcgcttcgat  ctctgtaacg  acttcgcccg  ggaggtcccg
1381 gtgctggcgc  tcggtaccgc  gcccgcactc  aagggcgtgg  accccgaccg  tctccgggtc
1441 tggacctcgg  cgacccgggt  atgcctggac  gccaggtca  gcccgcaaca  gctcgcgggt
1501 accgaacagg  cgctgaccgc  cctcgacgag  atcgacgcgg  tcaccggcgg  tcgggacgcc
1561 gcggtgctgg  tgggggtggg  ggccgagctg  gcggccaaca  cgggtgggaa  cgcgtcctg
1621 gccgtcaccg  agcttcccga  actggcggca  cgacttgccg  acgaccggga  gaccgcgacc
1681 cgtgtggtga  cggaggtgtc  gcggacgagt  cccggcgctc  acctggaacg  ccgcaccgcc
1741 gcgtcggacc  gccgggtggg  cggggtcgac  gtcccgaccg  gtggcgaggt  gacagtggtc
1801 gtcgccgcgg  cgaaccgtga  tcccagagtc  ttcaccgatc  ccgaccggtt  cgacgtggac
1861 cgtggcggcg  acgccgagat  cctgtcgtcc  cggcccggct  cgccccgcac  cgacctcgac
1921 gccctgggtg  ccacctggc  cacggcggcg  ctgcggggcg  ccgcgccggg  gttgccccgg
1981 ctgtcccgtt  ccgggcccgt  gatcagacga  cgteggtcac  ccgtcgcccg  tggcttcagc
2041 cgttgcccgg  tcgagctgta  gaggaagaac  gatgcgcgtc  gtgttttcat  cgatggctgt
2101 caacagccat  ctgttcgggc  tggteccgct  cgcaagcgcc  ttccaggcgg  ccggacacga
2161 ggtacgggtc  gtcgcctcgc  cggccctgac  cgacgacgtc  accggtgccc  gtctgaccgc
2221 cgtgcccgtc  ggtgacgacg  tggaacttgt  ggagtggcac  gccacgcgg  gccaggacat
2281 cgtcgagtac  atgcggaccc  tcgactgggt  cgaccagagc  cacaccacca  tgtcctggga
2341 cgacctcctg  ggcattgcga  ccaccttcac  cccgaccttc  ttcgccctga  tgagccccga
2401 ctgctcctc  gacgggatgg  tcgagttctg  ccgtcctctg  cgteccgact  ggatcgtctg
2461 ggagccgctg  accttcgccc  ccccgatcgc  ggcccgggtc  accggaacct  cgcacgcccg
2521 gatgctgtgg  ggtccggacg  tcgccaccgg  ggcccggcag  agcttctctg  gactgctggc
2581 ccaccaggag  gtggagcacc  gggaggatcc  gctggccgag  tggttcgact  ggacgctgcg
2641 gcgcttcggc  gacgaccgc  acctgagctt  cgacgaggaa  ctggtgctgg  ggcagtggac
2701 cgtggacccc  atccccgagc  cgctgcggat  cgacaccggc  gtccggacgg  tgggcatgcg
2761 gtacgtcccc  tacaacggcc  cctcgggtgt  gccgcctgg  ctgttgccgg  aaccggaacg
2821 tcggcgggtc  tgcctgaccc  tcggcgggtc  cagccgggaa  cacggcatcg  ggcaggcttc
2881 catcggcgag  atgttgagcg  ccatacgcca  catcgacgcc  gagttcgtgg  ccaccttcga
  
```



2941 cgaccagcag ttggtcggcg tgggcagcgt tccggcaaac gtccgtaccg ccgggttcgt  
3001 gccgatgaac gtcctgctgc ccacctgcgc ggccaccgtg caccacggcg gcaccggcag  
3061 ttggttgacc gccgccatcc acggcgctacc gcagatcatc ctctcggacg ccgacaccga  
3121 ggtgcacgcc aagcagctcc aggacctcgg cgcggggctg tcgctcccgg tcgcggggat  
3181 gaccgccgag cacctgcgtg gggcgatcga gcgggttctc gacgagccgg cgtaccgcct  
3241 cgggtgcggag cggatgcggg acgggatgcg gaccgaccgg tcgccggccc aggtggtcgg  
3301 catctgtcag gacctggccg ccgaccgggc ggacgcgggc aggcagccgc gtcgaaccgc  
3361 cgagccgcac ctgccgcgat gacttccacc accaccggga ccggctgatg ccgggtcccgg  
3421 aatccacacg ccgactttcc ttctgacacg agggggcccc ggtggttacc tccaccaact  
3481 tggacacgac agcacggccg gcactgaact cgttgaccgg gatgcgggtc gtcgccgcct  
3541 tcctggtctt cttcacgcac gtcctgtcga ggctcatccc gaacagctac gtgtacgccg  
3601 acggcctgga cgccttctgg cagaccaccg gacgggtggg ggtgtcgttc ttctttatc  
3661 tcagcgggtt cgtgctgacc tgggtcggcg gggccagcga ctccggtgtg tcgttctggc  
3721 gcagacgggt ctgcaagctc tcccccaacc acctggtcac cgccttcgcc gccgtgggtg  
3781 tgttccctgg caccggggcag gcgggtgagc gtgaggcgct gatcccgaac ctctgctga  
3841 tccacgcctg gttcccggcc ctggagatct ccttcggcat caaccgggtg agctggctgt  
3901 tggcctgcga ggcgttcttc tacctgtgct tcccgtgtt cctgttctgg atctccggta  
3961 tccgcccgga gcggctgtgg gcctggggcc cgtggtgtt cgcgcgac tcggcggtac  
4021 cgggtggtcg cgacctctg ctgccagatt ccccgcgcgt gatcccgggg cttagtact  
4081 ccgccatcca ggactgggtc ctctacacct tccctgcgac gcggagcctg gagttcatcc  
4141 tcgggatcat cctggcccg c atcctgatca ccggtcgggt gatcaacgtc gggctgctcc  
4201 ccgcgggtgt gttgttccc gtcttctctc tcgctcgtc cttcctgccg ggtgtctacg  
4261 ccatctcttc gtcgatgatg atccttcccc tgggtctgat catcgccagc ggcgcgacgg  
4321 ccgacctcca gcagaagcgc accttcatgc gtaaccgggt gatggtgtgg ctccggcagc  
4381 tctccttcgc gctctacatg gtccacttcc tgggtatcgt ctacggggcg gacctgctgg  
4441 ggttcagcca gaccgaggac gcccgcgtgg gtctcgcact cttcatgatc attccgttcc  
4501 tcgcgggtct cctggtgctg tcgtggctgc tgtacagggt cgtcgagcta cccgtcatgc  
4561 gtaactgggc ccgcccggcc tccgcccggc gcaaaccgcg cacggaaccc gaacagaccc  
4621 cttcccggcg gtaagaagga cgggtgcacg gtgaccacct acgtctggtc ctatctgttg  
4681 gagtacgaga gggaaacgag cgacatcttc gatgcgggtc agaaggctct cgcagtggtc  
4741 agcctgatcc tcggtcagag tgtggagaac ttcgagaccg agtacgcccg ctaccacggg  
4801 atcgcgcaat gcgtgggctg cgacaacggc accaacgctg tgaaactcgc gctggagtcg  
4861 gtaggtgtcg gacgcgacga cgaggtcgtc acggtctcca acaccgccc cccacagtc  
4921 ctggccatcg acgagatcgg cgcgcggccg gtcttctgtg acgtccgcga cgaggactac  
4981 ctcatggaca ccgacctggt ggaggcggcg gtaccccgcg gtaccaaggc catcgtcccg  
5041 gtgcacctgt acgggcagtg cgtggacatg acagccctgc gggaaactgg cgaccggcgg  
5101 ggcctcaagc tcgtggagga ctgcgcccag gccacagggt cccggcgggg cggtcggctg  
5161 gccgggacga tgagcgacgc ggcggccttc tcgttctacc cgacgaaggc cctcggcgcc  
5221 tacggcgacg gcggcgcggt cgtcaccaac gacgacgaga cagcccgcgc cctgcgacgg  
5281 ctgcgggtact acgggatgga ggaggtctac tacgtacccc ggaccccggg tcacaacagc  
5341 cgcctcgacg aggtgcaggc cgagatctct cggcgcaaac tgacccggct cgacgcgtac  
5401 gtcgcgggtc ggcgggcggt cgcccagcgg tacgtcgacg ggctcgccga cctccaagac  
5461 tcgcacggcc tcgaactccc agtggtaacc gacggcaacg aacacgtctt ctacgtgtac  
5521 gtcgtccgcc acccgcgccg cgacgagatc atcaagcgtc tccgggacgg gtacgacatc  
5581 tccctgaaca tcagctaccc ctggccgggt cacacctatg ccggcttcgc ccacctcgg  
5641 gtcgcgtcgg ggtcgtgccc ggtcacgaa cggttggccg gcgagatctt ctcccttccc  
5701 atgtacccct ccctccctca cgacctgcag gacagggtga tcgaggcggg gcgggaggtc  
5761 atcaccgggc tgtgacgagc ccgcgtgtcg tcagcgaaga cccactctgg aaggccgggt  
5821 catgccgaac agccactcga ccacgtcgag caccgacgtc gcccgcgtac agcgggcgga  
5881 catctaccac gacttctacc acggccgtgg caagggatac cgtgccgaag ccgacgcgt  
5941 cgtggagggt gcccgcaagc acacccaca ggcggcgacc ctgctggacg tggcctgcgg  
6001 gaccggatcc cacctggtcg agctggcgga cagcttccgg gaggtggtgg gggtcgacct  
6061 gtcggccgcc atgctcgcca ccgcccggcg caacgacccc gggcggggaa tgcaccaggg  
6121 cgacatgcgc gacttctccc tcgaccgcag gttcgacgtc gtcacctga tgttcagctc  
6181 caccggttac ctctcgcacg aggcggaact ggaccgtgcc gtggcgaacc tggccgggtc  
6241 cctcgcgcct ggcggcaccc tcgtcgtgga gccctgggtg ttcccggaga cgttccggcc  
6301 cggttgggtc ggggcccagc tggtcaccag cggtgaccgg aggatctccc ggatgtcgca  
6361 caccgtcccg gcgggtctgc ccgaccgcac cgcctcccgg atgaccatcc actacacggt  
6421 ggggtcaccg gaggcgggga tcgagcactt caccgaggtg cacgtgatga cctgttctgc  
6481 ccgcgcccgc tacgagcagg ccttccagcg ggcgggctg agctgctcgt acgtcggcca  
6541 cgacctgttc tcgccggggc tttctcgtcg ggtcgccgcg gagccggggc ggtgagggtc  
6601 gaggagctgg gcacgagggg ggtcttcacc ttcaccccgc agacgttcgc cgacgagcgg  
6661 ggggtgttcg gcacggcgta ccaggaggac gtgttctgtg cggcgctcgg ccgcccgtg  
6721 tcccgggtgg cccaggtcag caccaccggg tcccggcggg gtgtgggtcc ggggggtgcac

6781 ttcacgacga tgcccggctc catggcggaag tacgtctact gcgccagggg tagggcgatg  
6841 gacttcgccc tcgacatccc gcccggttcc ccgaccttcg gccggggccga gccggtcgag  
6901 ctctccgccc agtcgatggg cgggctgtac cttcccgtgg gcatggggcca cctgttcgtc  
6961 tccctggagg acgacaccac cctcgtctac ctgatgtccg ccggttacgt ccccgacaag  
7021 gaacggggcg tgcacccctt ggatccggag ctggcggttg cgatcccggc cgacctcgac  
7081 ctcgtcatgt ccgagcggga cccgggtcgca cccaccttcg gggaggcccg ggaccagggg  
7141 atcctgcccc actacgccgc ctggccggcc gccgcgcacc ggggtggtgc gacgtgaccc  
7201 cggccggggc tgccggccgg tgggtggtgct cggcgcgctc ggtgcggttc gtcgcccggc gggaggtcgt  
7261 caccacgccc ctggccgacc tcccgggtgc ggtgcggttc gacgcaccgg gtggacctca ccgaaccggg  
7321 cgtgcccttc ggtgccgtcg ccgactacga gacgcaccgg gtggacctca ccgaaccggg  
7381 agcgctcgcg gaggtggtcg ccgacgcccg ggcgggtctc ccgttcgccc cccagatcag  
7441 ggggtacgtca ggggtggcgga tcagcgagga cgacgtggtc gccgaacgga cgaacgtcgg  
7501 cctgggtccgg gacctgatcg ccgtcctgtc ccgtcgcggc caccgcccgg tgggtggtcct  
7561 cccggggcagc aacacgcagg tcggcagggg caccgcccgg cgggtcatcg acggcagcga  
7621 gcaggaccac cccgagggcg tctacgacag gcagaaacac accgggggaa agctgctcaa  
7681 ggaggccact gcggccgggg cgatccgggg gaccagtctg ccggtgcccc cgggtgttcgg  
7741 ggtgcccggc gccggcaccg ccgacgaccg ggggggtggtc tccaccatga tccgtcgggc  
7801 cctgaccggc caaccgctga cgatgtggca cgacggcacc gtccggcggtg aactgctgta  
7861 cgtgaccgac gccgcccggg ccttcgtcac cgccctggac caccgcccag cgtcgcggg  
7921 acgccacttc ctggtgggga cggggcggtc ctggccgctg ggcgaggtct tccaggcggt  
7981 ctcgcgacgc gtcgcccggc acaccggcga ggaccgggtg ccggtggtct cgggtgccgc  
8041 tccggcgcac atggaccgt cgacacctgc cagcggtgag gtcgacccc cccgggtcac  
8101 ggctgtcacc ggggtggcggg ccacgggtcac gatggcgag gcggtcgacc ggacgggtggc  
8161 ggcgttgggc ccccgccggg ccgcccggcc gtccgagccc tctgaccgg gggtaccggg  
8221 gttcgtccta cggcacccgg ccgtcgacgg ccggtgcccg gaagatcgct tcgagttccc  
8281 ggagttcctc ctgcgccagc gtcagctcgg ccggtgcccg ccggtgcccg gaagatcgct tcgagttccc  
8341 gtgtgcgggg gccgatgaca gcgccaggga tcccggggcg ggacaggacc caggccagac  
8401 cgacctcggc cgggtccgcg ccgaggcgct ggcagtagtc ctgtagcgc tcgacgaggg  
8461 ggcgtacggc ggggaggagc acctggggcg gtccctgcgc cgacttgacg gcggttccgg  
8521 ctgccaactt ctccagtacg ccgtcgagca gcccgccgtg cagggggggac caggcgaaca  
8581 cgcccccccc gtacgcctgg gcggcgggca ggacgtccag ctcggggtgg cggacggcca  
8641 ggttgtacag gcaactggtg gagatcatgc cgagcaggtt gcggcggtgc gcgctctcct  
8701 gggcgggcgg gatgtgccag cccgccaggg tggaggagcc gacgtaccgg accttcccac  
8761 tgccgaccag atgttcggcg gcctgccaca cctcgtccca cgggtgcccg cggtcgatgt  
8821 ggtgcgtctg gtagatgtcg atgtggtcga ccccgaggcg gcggagggag ttctcgagg  
8881 cggcgacgat gtgtcggcg gagagcccg gcgtcgttgac ccgttcgctc atctcgctgc  
8941 ccaccttggg cgccaggacg gtctcctcgc gtcgacctcc gccctggggc aaccaccgtc  
9001 cgacgagttc ctcggtgtgg cccttgtaga gccgcccagg gtagatgtcg gcggtgtcga  
9061 tgcagttgac gccccgctcg agggcggtgg ccatcagccg cagcgcgctc tctcggtca  
9121 cccgtccact gaagttcacg gtgcccagcc agagtcggct ggtgtgcaac gccgatcgtc  
9181 cgacgcgtac ccgggcggac ccggccccgg tgggttccac gtcggtcacc tgtcggcgcg  
9241 gtgctggtgg gcgagcgct ccagcacggg tacgacctcg gcgggggtcg gcgcccag  
9301 cgcctcctgc cgcagcttct cggcgctctc ggcgtgggaa ccgtcctcga ccactgtggc  
9361 gagagcctgc cagagggtgt cggcgctcga ctcgtccgga cggaggaga caccgctcc  
9421 cagctcggcg gtgcgctgac cacgcaggac acagtccac tcgtgggcca cggagatctg  
9481 cggtagcccg tgggtgcagc cgggtggcca gcttcgggca ccgcccgtgg ggatgacggc  
9541 ggcacagccc ggcagcagga tgttcatggg aacgaagtcc accaggcgga cgttgtccgg  
9601 caccgacgcc ggatcgagcc cggagcgggt caccacgata tcgcccgtcg accgcccag  
9661 ggtggccagt gtccggagga actcctgcgg gttcgaggtg atgccagcg ccgagtatcc  
9721 cccgggtgaag cagaccggc ggactccgtc cgaggtcctg agccactgcg gcacgacgga  
9781 ggacccggtg tagggcaaa tcccgggtgt caccgactcc agtccgggtc ccaggcgga  
9841 gctctcgggc agctggtcga cgctccactg tccgacagcg aggtcctcgc tgtagtcgag  
9901 gccgaaccgg ccggcgacct cgggtgagcca gccgcccagc ggggtccggc ggtcgtcggc  
9961 gggacgctgc ccgcccaggt cctgggagcg gctgcggaag tagccgggtg ggtcgtcggc  
10021 ccacagcagc cgggcgtggg cggccccgca ggccttggcc gcgaccgcc cggcgaagg  
10081 gaagggtccc cagagcacca ggtcgggacg ccagtcctat gcgaactcga cgagttcgtc  
10141 gacgaaggag tcgttgttga ccaccgggaa gacgaaccgg gaggtggcct cctcgatgcc  
10201 gtgcaggaa tcccacgagc gcagttccgg tccgcgtcgg gcgaagtcca ggtcgggtg  
10261 gtagcgggtg acctgcgcgg cggcctcagg ggagatgtcg aagagtcggg ggtccgagcc  
10321 gagtggcacc gaggtcagtc ccgcccagac gacgacgtcg gtgagctcgg gctgactggc  
10381 caccgggacg tcgtggccgg cgggtgtcag cggcccaggc agggggagca ggccctggaa  
10441 gtgggtacgg tgcgcgaac aggtgagcag gacccgact ggtcactcct tggtcgagat  
10501 gagggcgga acggtccggc cgatgccctc ggccagcggc acccgggggg gccagccgg  
10561 cagcgtccgg aactcgggtg agtcgaagtc gtcgctgcgg aagtcgttgg cctcggcggt



10621 ctccggtgga gggacgctga cgacgggac cgcagggttg ccggtctgac gtgccacgct  
10681 ggcggcgacg gtctcgaaga tctcgccgag gggtcggggc tcgtccgcgc tcggcgcca  
10741 gacgtcgccg accagcgccct cgtggttgtg cagtgcggcg gtgaacgcgg tggccacgtc  
10801 ctcgacgtgc aggaggttgc ggcgcacgct gccctcgtgc cacatcgtga tcggctcacc  
10861 ggcgagggct cgcgggatca tggcgggtgac gacaccccg ccggtctgcc ccgacgggccc  
10921 gctgtggccg tagatcgccg gcaggcgag gatcaccccg tcgacgaccc cgtcctcggg  
10981 ggcctgacgc aggatccgct cggcctcgat cttgtgctgg gcgtaccggc tggggggcggc  
11041 ggggttcgcg gcctgggtgg tgctggcgaa caggagcacc ggcgcgggtc cgggtcttgc  
11101 ccgcagcgcg gcgacgaggt cgcgcacgat gcccggttg acgcgttcgg cctcgggac  
11161 cgtggcgcg ctgcccagg tgcacccgcc ggcggcgtag gcgaccagat gcacgacgac  
11221 gtcggtgtcg gcgacgacct gcgcgacccg gccgggttcg agcaggtcga ctcaagggtg  
11281 ctcgatcccg gcgctgcctg gtggctgggtc gcgagacccg gtgcgcgcga cggcccgag  
11341 tcggagaggg tgtgtggtaa attcgcgaag aaggcgctt ccgacgaatc cagaaacgcc  
11401 gagaagtgtg acatgtcttg tcatctacta atgcattccg atagccaccg gcgcattgaa  
11461 tccatttgtt ccccccaggg tgggtgcggg tgacaaatcc ggcctcaggc cggcctcaag  
11521 cctctttcga gcgggtgctg aggttcccg cgtaccctcg gtggcctgcg ttcgggcggg  
11581 tgtcggggaa agggcggtac gaggagttcg gtaggcgctc gcggcgcgta ctccgggact  
11641 gatccgggtc gacgccccga cgcgtgacag ggcgtcgatc cgtgccgccc gtaccgcccg  
11701 ttttcggcga tggtcgcaga ttcctcccga cgtggtggac tcattggttc tcccggtgt  
11761 ggcgcacccg tcgggtggcct cgtcgggggt gtcggagacc gggtcgatcg ccgtcccgg  
11821 ccgtgccgac cagggtcggg ccgtcgccga ggtgggtcac cgtcgggtgg acccggtccg  
11881 ccggcgcca ccgccgatc gtgcccacct tcgcctccgc gggtaaagtc ttcgtcgatc  
11941 tgatcgacac ttccggcgac gctatcaccg gagcattccc cggcaccacc ggtcgatgcc  
12001 tcgcgctttc caaacaggga aaacagcagc tcacagcggt tccaggcgcc gggcaatcct  
12061 agcgaagagt ctcgatgggg tcaaggtgaa ttctgtcaca gatgtttttg ttaaatgtac  
12121 tttcttcagc caccctcgac gttcatacaa ttggccggca tctctaccaa gggggagtga  
12181 gtggttgacg tgcccgatct actcggcacc cggactccgc acccagggcc gctcccattc  
12241 ccgtggcccc tgtgcggtca caacgaaccg gagctgcggg cccgcgcccc tcaattgcac  
12301 gcatatctcg aaggcatttc cgaggatgac gtggtggccg tcggcgccgc cctcgcgcgc  
12361 gagacacgcg cgcaggacgg gccgcaccgc gccgtcgtcg tggcctcctc ggtcaccgag  
12421 ctgaccgccc cgtcgcgcgc cctcgcccag ggcgcgccac acccctcggg ggtacgcggg  
12481 gtcgcccgac ccacggcacc ggtggtgttc gtcctgcccg gtcaggcgcc ccagtggccc  
12541 ggcattggcga cccgactgct cgcgcagtcg cccgtcttcg ccgcggcgat gcgggcccgc  
12601 gagcgggccc tcgacgaggt caccgactgg tcgttgaccg aggtcctgga ctaccccag  
12661 cacctgcgcc gcgtcgaggt ggtccagccc gcgctcttcg ccgtgcagac ctactggcc  
12721 gccctgtggc ggtcgttcgg ggtgcgaccc gacgcccgtac tcggacacag catcggtgag  
12781 ctggccgccc ccgaggtctg cggcgccgctc gacgtcgagg ccgcgcgcgc ggcgcgccgc  
12841 ctgtggagcc gcgagatggt cccactggtg ggccgggggtg acatggcgcc ggtggcgctc  
12901 tccccggccg agctggcagc ccgggtcgag ccgtgggacg acgacgtcgt gccggccggg  
12961 gtcaacggtc cccggtcggg gctgctcacc ggcgctcccg agcccatcgc acggcggggtc  
13021 gccgagctgg cggcacaggg cgtacgcgcc caggctcgta acgtgtcgat ggcggcgac  
13081 tcggcgcagg tcgacgcccgt cgcgcagggc atgcgctcgg cgtgacctg gttcgcccc  
13141 ggcgactccg acgtgcccta ctacgcccgc ctcaccggcg ggcggctgga caccgggaa  
13201 ctccggcgcc accactggcc gcgcagtttc cggctcccgg tgcgcttcga cgaggcgacc  
13261 cgtgcgggtc tggaaactga gcccggcacg ttcacagagt cgagcccga cccggtgctg  
13321 gcggcctccc tgcagcagac cctcgacgag gtcgggtccc cggccgcgat cgtgccgacc  
13381 ctgcaacgcg accaggcgcg tctgcggcgg ttcctgctcg ccgtggcgca ggcgtacacc  
13441 ggtggcggtga cagtcgactg gaccgcccgc taccocgggg tgacccccgg ccacctgccg  
13501 tcggccgctc ccgtcgagac cgacgaggga ccctcgacgg agttcgactg ggcgcgccc  
13561 gaccacgtac tgcgcgcgcg gctgctggag atcgtcggcg ccgagacggc cgcgctcgcc  
13621 gggcgggagg tcgacgcccc ggccaccttc cgggaactgg gcctcgactc ggtcctcgcg  
13681 gtgcagctgc ggacccgcct cgccacggcg accggcgggg atctgcacat cgccatgctc  
13741 tacgaccacc cgaccccgca cgccctcacc gaggcgctgc tgcgcggccc gcaggaggag  
13801 ccggggcggg gtgaggagac ggcacacccc acggaggcgg aaccgacga acccgtcgcc  
13861 gtggtcgcca tggcgtgccg gctgcccggc ggcgtcacct caccggagga gttctgggag  
13921 ctgctggccg agggcgggga cgcgctcgcc gggctgccc cccgaccggg atgggacctg  
13981 gactcgctgt tccaccggga cccgacccgg tcgggcacgg cgcaccagcg cgtggtggc  
14041 ttcctcaccg gcgccacctc cttcgacgct gccttcttcg ggctgtcgcc acgggaggca  
14101 ctggccgctc agccgcagca gcggatcacg ttggagctgt cgtgggaggt gctggaacgc  
14161 gccgggatcc ccccgacgtc gttgcggacc tcccggaccg ggggtgtcgt cggctgatc  
14221 ccccgaggat acggcccccg gctggccgag ggggggtgag gcgtcgaggg ctacctgatg  
14281 accgggacca ccaccagcgt cgcctccggt cgggtcgcc acaccctcg cctggagggg  
14341 ccggcgatca gcgtcgacac cgcctgctcg tcgtcgctcg tcgcccgtga cctggcgctg  
14401 cagtcgctgc ggcgcggcga gtcgacgat gcgctcgccc gtggcgtgac ggtgatgccc

14461 acaccgggca tgctcgtgga cttcagtcgg atgaactccc tcgcccccca cggacgggtcc  
14521 aaggcggttct cggccgcccgc cgacgggttc ggcattggccg aaggcgccagg gatgctcctg  
14581 ctggaacggc tctcggacgc ccgcccgcac ggccaccggt tgctcggcgt gatcaggggc  
14641 accgctgtca actccgacgg cgcgagcaac ggactctccg ccccgaaagg ccggggcccag  
14701 gtccgggtga tccgacaggc cctcggccgag tccgggctga cggccccacac cgtcgcagtc  
14761 gtggagaccc acggcaccgg caccgcctc ggtgatccga tcgaggcacg ggcgtctctc  
14821 gacgcgtacg gcggtgaccg tgagcaccgc ctgcggtacg gctcgggtcaa gtccaaacatc  
14881 gggcacaccc agggccgccc cgggtgtcgc ggtctgatca aactgggtgtt ggcgatgcag  
14941 gccgggtgtcc tgccccgcac cctgcacgce gacgagccgt caccggagat cgactgggtcc  
15001 tcggggcgga tcagcctgct ccaggagccc gctgcctggc ccgcccggcga gcggccccgc  
15061 cgggcccgggg tgcctcgtt cggcatcagc ggcaccaacg cacacgcgat catcgaggag  
15121 gcgcccgcga ccggtgacga caccgaccc gaccggatgg gcccggtggg gccctgggtg  
15181 ctctcggcga gcaccggcga ggcgttgccg gcccgggcg cgcggctggc cgggcacctc  
15241 cgcgagcacc ccgaccagga cctggacgac gtcgcctact cgctggccac cggtcggggc  
15301 gcgctggcgt accgtagtgg gttcgtgccc gccgacgcgt ccacggcgct gcggatcctc  
15361 gacgaactcg ccgcccgggtg atccggggac gcggtgaccg gcaccgcccg cggcccgag  
15421 cgcgtcgtct tcgtcttccc cggccaggga tggcagtggg cggggatggc agtcgacctg  
15481 ctcgacggcg acccggtctt cgcctcgggt ctgcccgggt gcgcccagc gttggaaccg  
15541 tacctggact tcgagatcgt cccgttctct cgggcccagg cgcagcgccg gacccccgac  
15601 cacacgctct ccaccgaccg cgtcgcagtg gtccagccgg tgctgttcgc ggtgatgggtg  
15661 tccctggcgg cccgggtggcg ggcgtacggg gtggaaccgg cggccgtcat cggacactcc  
15721 cagggggaga ttgcccggc gtgtgtggcc ggggcgtctc cgctggacga cgcggcccgg  
15781 gcggtggccc tgcgcagccg ggtcctcgcc accatgcccg gcaacggcgc gatggcctcg  
15841 atcgccgctt ccgtcgacga ggtggcgccc cggatcgacg ggccgggtcga gatcgccgc  
15901 gtcaacgggtc cgcgcgcggt ggtgggtctc ggccgaccgt acgacctgga ccgcctggtc  
15961 gcctcctgca ccgtcgaggg ggtgcccggc aagcggtgct cgggtggacta cgcgtcgac  
16021 tcctcgcacg tcgaggccgt ccgtgacgcg ctccacgccc aactcggcga gttccggccg  
16081 ctgcccgggt tcgtgcccgt ctactcgaca gtcaccggcc gctgggtcga gcccgcgaa  
16141 ctgacgcccg ggtactgggt tcgcaacctg cggccacagg tccgggttcgc cgacgcgggtc  
16201 cgtccctcgc ccgaccaggg gtacacgacg ttccctggagg tcagcggcca cccgggtgctc  
16261 accacggcga tcgaggagat cgggtgaggac cgtggcggtg acctcgtcgc tgtccactcg  
16321 ctgacgacgt gggccggcgg tcccgctcgc ttccggtccg cgctggcccc cgccttcgtg  
16381 gccggcgctc cagtggactg ggagtcggcg taccagggtg ccggggcgcg tcgggtgccc  
16441 ctgcccacgt acccggtcca gcgtgagcgc ttctgggttg cgggtaccga tcgaatggca cccgaccgat  
16501 gtcgcccact ccgacgacgt ctcgtccctg cgggtaccga tcgaatggca cccgaccgat  
16561 ccgggtgagc cgggacgggt cgacggcacc tggctgctgg cgacgtacc cggtcggggc  
16621 gacgaccggg tcgaggcggc gcggcaggcg ctggagtccg ccggggcgcg ggtcgaggac  
16681 ctggtgggtg agccccggac gggccgggtc gacctggtgc ggccggtcga cgcgtgggt  
16741 ccgggtggcg gcgtgctctg cctgttcgct gtcgcccagc cggccggcga acactccccg  
16801 ctggcggtga cgtcgttgtc ggacacgctc gacctgacct aggcgggtgg cgggtcgggc  
16861 cgggagtgtc cgatctgggt ggtcaccgag aacgcggtcg ccgtcggggc ctccgaacgg  
16921 ctccgcgacc cggcccacgg cgcgctctgg gccctcgggt gggtcgtcgc cctggagaac  
16981 cccgcggtct ggggcggcct ggtcgacgtg ccgtcgggtt cggtcgcccga gctgtcgcgt  
17041 cacctcggga cgacctgtc cggcgccggc gaggaccagg tcgcccctcg acccgacggg  
17101 acgtacgccc gccgggtggg cagggcgggc gcgggcggca cgggcccgtg gcagccccgg  
17161 ggcacgggtg tcgtcaccgg cggcaccggc ggggtcgggt ggcacgtcgc ccgggtggctg  
17221 gcccgccagg gcaccccggt cctgggtgct gccagccgce ggggaccgga cggcgacggg  
17281 gtcgaggagc tactcaccga actcgcggac ctgggcaccc gggccaccgt caccgcctgc  
17341 gacgtcaccg accgggagca gctccgtgcc ctctcgcga ccgtcgacga cgagcaccgg  
17401 ctgtcggcgg tgttccacgt cgccgcgacg ctcgacgacg gcaccgtcga gacctcacc  
17461 ggtgaccgca tcgaacgggc caaccggggc aagggtgctc gtgcccgcaa cctgcacgag  
17521 ctgacccggg acgcccacgt cgacgcgttc gtgctcttct cctcctccac cgcgcgttc  
17581 ggcgcgcccg ggctcggcgg ctacgtcccg ggcaacgcct acctcgacgg tctcggcccag  
17641 cagcgacgca gcgagggact cccggccacc tcgggtggcgt ggggtacctg ggcgggcagc  
17701 gggatggccg aggggtccgt cgcgcacggg ttccgcccgg acgggggtcat ggagatgcac  
17761 cccgaccagg ccgtcgaggg tctccgggtg gcaactgggt aggggtgagg agccccgatc  
17821 gtcgtcgaca tcagggtggga ccgggttctc ctgcggtaca ccgcgacgc cccaccccgg  
17881 ctcttcgaca ccctcgacga gggccgctcg gccgcgcccg gtcccgcgc cgggcccggg  
17941 gtggcggcgc tggccggggt gcccgctcgg gaacgcgaga aggcgggtct cgacctggtg  
18001 cggacgcacg cggctgcccgt cctcggccac gcctcggccc agcagggtgc cgtcgacagg  
18061 gccttcgccc aactcggcgt cgactcgtcg tcggccctgg aactgcgcaa ccggctgacc  
18121 actgcgacgg ggggtccggct ggccacgacg acgggtcttc accacccgga cgtacggacc  
18181 ctggccggac acctggccgc cgaactgggc ggcggatcgg ggcgggagcg gcccgggggc  
18241 gagggcccga cgggtggcccc gaccgacgag ccgacgcgca tcgtcgggat ggctgcccg

18301 ctgccggggg gagtggactc accggagcag ctgtgggagt tgatcgtctc cgggcggggac  
18361 accgcctcgg cggcaccgga ggaccggagc tgggatccgg cggagttgat ggtctccgac  
18421 acgacgggca cccgtaccgc ctccggcaac ttcattgccc gggcggggca gtccgacgag  
18481 gcgttcttcg ggatctcgcc gcgtgaggcg ttggcgatgg atccgcagca gcggcacgac  
18541 ctggagacca cctgggaggc gctggagaac gccggtatcc ggcccagatc gttgcgaggt  
18601 acggacaccg gtgtcttcgt gggcatgtcc catcaggggt acgccaccg cggcccgaag  
18661 cccgaggacg aggtcgacgg ctacctgttg acaggcaaca ccgcgagcgt cgcctccggt  
18721 cggatcgctg acgtgttggg gttggagggg ccggcgatca ctgtggacac ggctgttctg  
18781 tcgtcgcttg tggcggttga cgtggcgggc ggttcgttgc gttctgggga ctgtggtctg  
18841 gcgggtggcg gtgggggtgt ggtgatggcc ggtccggagg tgttcaggga gttctcccgg  
18901 cagggcgctg tggctccgga cggcagggtg aagcccttct cggacgaggg cgacggcttc  
18961 ggtctggggg aggggtcgcc ctccgtcggt ttgcagcggt tgtcgggtggc ggtgcgggag  
19021 gggcgctcgg tgttgggtgt ggtgggtggg tcggcggtga atcaggatgg ggcgagtaat  
19081 ggggtggcg cggcgctcgg ggtggcgag cagcggtga ttcggcgggc gttgggtcgt  
19141 gcgggtgtgt cgggtgggga tgtgggtgtg gtggaggcgc atgggacggg gacgcgggtg  
19201 ggggatccgg tggagttggg ggcgttcttg gggacgtatg ggggtgggtc ggggtggggtg  
19261 ggtccgggtg tgggtgggttc ggtgaaggcg aatgtgggtc atgtgcaggc ggcggcggtg  
19321 gtgggtgggtg tgatcaagggt ggtgttgggg ttgggtcggg ggttgggtggg tccgatgggtg  
19381 tgtcgggggtg ggttgtcggg gttggtggat tgggtcgtcg gtgggttggg ggtggcggtg  
19441 ggggtgcggg ggtggccggg ggggtgtgat ggggtgcgtc ggggtggggt gtcggcggtt  
19501 ggggtgtcgg ggacgaatgc tcatgtgttg gtggcgagg cggcggggtc ggtgggtggg  
19561 gcggaacggc cgggtggagg gtcgtcgcg ggggtgggtg ggggtgggtg tgggtgtgggtg  
19621 ccgggtgtgc tgtcggcaaa gaccgaaacc gccctgcacg cccaggcacg tccactcgcc  
19681 gaccacctgg agacgcaccc cgacgtcccg atgaccgac tgggtgtggac gctgacgcag  
19741 gcccgcacac gcttcgacag gcgcgcgggt ctcctcgccg ccgaccggac ccaggccgtg  
19801 gaacggctgc gcggcctcgc cggggggcga ccggggaccg gtgtgggtgt ggggggtggc  
19861 tcgggtgggtg gtgtgggtgt tgtttttctt ggtcagggtg gtcagtgggt ggggatggcg  
19921 cgggggttgt tgtcgggttc ggtgtttgtg ggtcgggtg tggagtgtga tgcgggtggg  
19981 tcgtcgggtg tgggggtttc ggtgttgggg gtgttggagg gtcgggtcgg tgcggcgctg  
20041 ttggatcggg tggatgtggg gcagccgggt ttgttcgtgg tgatgggtgt gttggcgcg  
20101 ttgtggcggt ggtgtggggg tgtgcctcgc gcgggtgggt gtcatcgcga gggggagatc  
20161 gcggcgcgcg tgggtggcggg ggtgttgtcg gtgggtgatg gtgcgcgggt ggtggcggtg  
20221 cgggcgcggg cgttgcgggg gttggccggc caccggcgga tggcctcgg acggcgaggc  
20281 cgcgacgacg tacagaagct cctcgacagc ggccctcggg cgggggaagct ggagatcgcc  
20341 gcgggtcaacg gccccgacgc ggtgggtggc tccggcgacc cccgagccgt gaccgagctg  
20401 gtcgagcact gtgacgggat cgggggtcgg gcccggaaga tcccgtcga ctacgcctcc  
20461 cactccgcac aggtcgagtc gctccgggag gagctgtctt ccgtcctggc cgggatcgag  
20521 ggccgcccgg cgacgggtgc gttctactcc accctcaccg gtgggttcgt cgacggcacc  
20581 gaactggacg ccgactactg gtaccgcaac ctgcgcacc cgggtgcgggt ccacgcgcgc  
20641 gtcgagggcg tggcagcgcg tgacctcacc acgttcgtcg aggtcagccc gcaccccggtg  
20701 ctgtcgatgg cgggtcgggga gacgcttgcg gacgtggagt ccgcccgtac tgtgggcacc  
20761 ctggaacgcg acaccgacga cgtcgagcgc ttcctcacct ccctcgccga ggcgcacgtc  
20821 caccgggtac ccgtggactg ggccggcggt ctcgggtccg gaacctgggt cgacctgccc  
20881 acctatccct tccagggacg gcgggttctg ctgcaccccg accgtgggtc gcgtgacgat  
20941 gtcgcccact ggttccaccg ggtcgactgg acggcgacgg ccaccgacgg gtcggcccga  
21001 ctcgacgggt gctgggtggg ggtcgtaacc gaggggtaca cggacgacgg ctgggtcgtg  
21061 gaggtgcggg ccgcctcgc cgcgggtggg gccgagccgg tgggtgacgac ggtcgaggag  
21121 gtcaccgacc ggggtcgggtga cagcgacgcg gtgggtgtcga tggctcgggt ggccgacgac  
21181 ggtgcggccg agacctggc gctgtgtcga cgtcgcagc cacaggcgct caccacccca  
21241 ctgtgggtgg tcaccgtggg ggcgtcgc cccgcccgtc cgggtgcagc ccccgaaacg  
21301 gcgacgggtg ggggggttgg ccttgtcgc tccctggaac gcggacaccg gtggaccggc  
21361 ctgctggatc tgccgcagac accggacccg cagctacgac cccggctggg cgaggcgctc  
21421 gccgggtgcg aggaccagg agcgggtcgc gccgacggc tacacggccg tcggatcgtc  
21481 cccaccccg taccggagc cgggcccgt accgcccgg gcgggacgat cctcgtcacc  
21541 gggggcaccg ccgggtctgg tgcgtcacc gcccgatggc tcgcccagcg cgggtgcccga  
21601 cactcgcgcc tgggtcagcg gcgcggggcg ggcaccgccc gcgtcgacga ggtgggtcgg  
21661 gacctgaccg ggctcggcgt acgggtgtcg gtgcactcct gcgacgtcgg cgaccgcgag  
21721 tcgggtcggc ccctgggtga ggagtgtaca gcagccgggt acgtgggtcg ggggggtggc  
21781 cagctgccc gtctgcccc gcaggtgcca ctgaccgaca tggacccggc cgacctcgcc  
21841 gacgtgggtg ccgtgaagg cgacggcgcg gtgcacctg ccgacctgtg cccggaggcc  
21901 gaactgttcc tgctgttctc ctccggggcc ggggtgtgg gcagtgcgg tcagggtgag  
21961 tacgcccgg gaaacgcct cctggacgcc ttcgcccga accggcgggg ccgggggtctg  
22021 cccgccacct cgggtggcgt gggggtctgg gcggccgggg ggatgacagg ggaccaggag  
22081 gcgggtgtcgt tctcgtgta gcggggcgta cggccgatgt cgggtgccgag ggcactggaa



22141 ggcgtggaac gggtcctcac cgccggggag accgcggtgg tcgtcgccga cgtcgactgg  
22201 ggggccttcg ccgagtcgta caccctccgc cgcccccggc cgctgctcca ccggctcgtc  
22261 acacctgcgg cggcgggtcgg cgagcgcgac gagcccgctg agcagaccct ccgggaccgg  
22321 ctggcgggccc tgccccgggc cgagcggtcg gcggagctgg tacgcctggt ccggcgggac  
22381 gccgcagccg tgctcggcag cgacgcgaag gccgtacccg ccaccacgcc gttcaaggac  
22441 ctccgggttcg actcgtctggc cgcggtccgg ttccgtaacc ggctggccgc ccacaccggt  
22501 ctgctgtctgc cggccaccct ggtcttcgag caccggaacg ccgcagccgt cgccgacctc  
22561 ctccacgacc gactcggcga ggccggcgag ccgacccccg tccggtcggt gggcgccgga  
22621 ctggccgcgc tggagcaggc cctgccccgac gcctccgaca cggagcgggt cgagctggtc  
22681 gagcgccctgg aacggatgct cgccggggtc cgccccgagg ccggagccgg ggccgacgcc  
22741 ccgaccgccg gtgacgacct gggggaggcc ggcgtcgacg aactcctcga cgcgctcga  
22801 cgggaactcg acgccagggtg aaccggaact gaccgcagcc gcagccgaag cagagaccga  
22861 ggacctgtga ctgacaacga caagggtggcg gagtacctcc gtcgtgcgac gctcgacctg  
22921 cgggcccggc gcaagcgccct gcgcgagctg caatccgacc cgatcgcggt cgtcggcatg  
22981 gcctgcggcc taccggggcg ggtgcacctc ccgcagcacc tgtgggacct cctgcgccag  
23041 gggcacgaga cgggtgtccac ctccccacc gggcgcggtt gggacctggc cgggctcttc  
23101 caccggacc cccgaccacc cgccaccagc tacgtcgacc ggggtgggtt cctcgacgac  
23161 gtggcgggct tcgacgccga gttcttcggg atctccccgc gcgaggccac ggccatggac  
23221 ccgcaacagc ggctgctggt ggagaccagt tgggagctgg tggagagcgc cggcatcgat  
23281 ccgcactccc tgcgtggcac cccgaccggc gtcttcctcg gcgtggcgcg gctcggctac  
23341 ggcgagaacg gcaccgaagc cggtgacgcc gagggctatt cggtgaccgg ggtggcacc  
23401 gctgtcgctt ccggggcggt ctctacgcc ctccgggtgg agggctcgtc gatcagcgtg  
23461 gacaccgcgt gctcgtcgtc gttggtggcg ctgcacctgg cggtcgagtc gctcgggctg  
23521 ggcgagtcga gtctcgtgtg cgtcggcggg gcggcggtca tggcgacacc aggggtgttc  
23581 gtcgacttca gccgccagcg ggcgttggcc gctgacggca ggtcgaaggc cttcggggcc  
23641 gccgccgacg ggttcggctt ctccgagggg gtctccctcg tctgtctcga acggctctcc  
23701 gagggcgaaa gcaacggcca cgaggtgttg gctgtcatcc gtggctccgc cctcaaccag  
23761 gacggggcca gcaacggtct cgccgcggcg aacgggaccg cccagcgcaa ggtgatccgg  
23821 caggcgctac gaaactgcgg cctgaccccg gccgacgtgg acgcccgtga ggcgcacggc  
23881 accggcacca cgtcggcga cccgatcgag gccaacgccc tgcggacac ctacggccgt  
23941 gaccgggatc cggaccacc gctgtggctg gggtcgggtga agtcgaacat cggccacacg  
24001 caggcgggcg cggggtcac cgggctgctc aagatgggtg tggcactgcy ccacgaggaa  
24061 ctgcccggca cctgcacgt cgacgagccc accccgcacg tggactggte ctcgggagcg  
24121 gtacgcctgg cgacccgggg ccggccgtgg cggcggggtg accggccgag gcgggccggg  
24181 gtgtcggcgt tcggcatcag cgggaccaac gcccacgtga tcgtcgagga ggcacccgag  
24241 cggaccaccg agcgcaccgt cgccggcgac gtcggcccgg tcccgtctgt ggtgtccgcc  
24301 cggtcggcgg cgccgctacg ggcccaggcg gccaggtcg ccgagctggt ggagggtcc  
24361 gacgtcgggc tggcgagggt cgggcgagc ctggccgtga cccggcgcg acacgagcac  
24421 cgggcgggcg tgggtggcgt gaccggggcc gaggcgggtg gggggctgcy cgaggtcgcg  
24481 gcggtcgaac cgcgcggcga ggacaccgtc accggggctg ccgagacgtc cggcgccacc  
24541 gtcgtcttcc tcttccccgg acaggggtcc cagtgggtcg ggatgggcgc ggagctgctg  
24601 gactcggcac cggcgttcgc cgacacgata cgcgcctgcy acgaggcgat ggcaccgttg  
24661 caggactggc cggctctccga cgtgctccgg caggagccgg gggcaccggg actggaccgg  
24721 gtcgacgtgg tgcagccggg gctgttcgcy gtgatgggtg cgttggcgcy gttgtggcag  
24781 tcgtacgggg tcacccccgc tgcgggtggg gggcactcgc agggggagat cgcgccgcc  
24841 cactgtggcg gtgcgtcttc cctcgccgac gcggcgaggc tgggtgggtg cggcagccgg  
24901 ttgctgcggc cgtgtccgg gggcgggcg atgagcgccg tcgcgtctcg tgaggccgag  
24961 gtacgccgcc gactgcggtc gtgggaggac cggatctccg tggccggcgt caacggaccc  
25021 cggtcgggtg tgggtggccg ggaaccggag gcgctgcggg agtggggagc ggagcgggag  
25081 gccgagggcg tacgggtccg cgagatcgac gtcgactacg cctcgactc gccgcagatc  
25141 gacaggggtc gtgacgaact cctgacggtc accggggaga tcgagccccg gtcggcgag  
25201 atcaccttct actcgacggt cgacgtccgt gctgtcgacg gcaccgacct ggacgcgggg  
25261 tactgggtacc gcaacctgcg ggagacggtc cggttcgccg acgcgatgac ccggttggcc  
25321 gactcgggat acgacgcgtt cgtcgaggtc agcccgcata cgggtgggtg gtcggcggtc  
25381 gccgagggcg tcgaggaggc aggtgtcgag gacgcctcg tctcgggcac cctgtcccg  
25441 ggcgacggcg gaccgggggc gttcctgcyg tcggcgccca ccgcccactg cggcggtgtg  
25501 gacgtcgact ggacgcccgc cctcccggga gctgcgacga tcccgttgcc gacgtacccg  
25561 ttccaacgga agccgtactg gctgcggtcg tctgtctccg cccccgctc ccacgatctc  
25621 gcctaccggg tgtcctggac gccgatcacc ccgcccgggg acggcgctact cgacggcgac  
25681 tggctgggtg tgcacccccg gggcagcacc ggatgggtcg acgggttggc ggcggcgatc  
25741 accgcccggc gtggccgggt cgtcgcccac ccgggtggact ccgtgacctc ccggaccggc  
25801 ctggccgagg cgtcgcggcg gcgggacggc acgttcgggg ggggtgctgtc gtgggtggcg  
25861 accgacgaac ggcacgtcga ggccgggtgc gtcgcccctg tgacctggc gcaggcgttg  
25921 ggtgacggcg gaatcgacgc accactgtgg tgcccgaccc agggaggcgt ccgtaccccc

25981 gtcgacgggtg acctggcccc accgggcgcag gccgccctgc acggtttctgc ccagggtcgcc  
26041 cggctggagc tggccccgcg cttcgggtggg gtgctcgacc tgccccccac cgtcgacgcc  
26101 gccgggacgc gtctggtcgc ggcgggtcctc gccggcgggcg gcgaggacgt cgtcgccgctc  
26161 cgtggcgacc gtctctacgg ccgtcgccctg gtcagggcgga ccctgccgcc gcccgggcggg  
26221 gggttcaccc cgcacggcac cgtcctgggtc accggcgcgcg ccgggtccggg gggcggtcgg  
26281 ctggccccggt ggctcgccga accggggtgcc acccgactcg tcctgccccg cgcacaccgg  
26341 ggcgaggagt tgctgaccgc gatccggggc gccgggtgcca ccgccgtggg gtgcaaccg  
26401 gaggcggagg cactgcgtac ggcgatcggc ggggagttgc cgaccgcgct cgtacacgcc  
26461 gagacgttga cgaacttcgc cggcgctgcc gacgccgacc ccgaggactt cgcgccacc  
26521 gtcgcggcga agaccgcgct gccgacggtc ctggcgaggg tgctcgggca ccaccgctc  
26581 gaacgggagg tctactgctc gtcgggtggc ggggtctggg gtggggtcgg catggccgcg  
26641 tacgccgcgc gcagcgccca cctcgacgcc ctggctcgag accgtcgcg cccggggcac  
26701 gccagcgctt cgggtggcctg gaccccggtg gccctgcccc gcgcggtcga cgacggtcgg  
26761 ctgcgcgagc gcggcctgcg cagcctcgac gtggccgacg ccctcgggac gtgggaacgt  
26821 ctgctccgcg ccgggtgcggg gtcgggtggc gtcgccgacg tcgactggtc ggtcttcaca  
26881 gagggttctc cggccatccg gccgaccccg ctcttcgacg aactcctcga ccggcgcggg  
26941 gaccccgacg gcgcgcccgt cgaccggccg ggggagccgg cgggcgagtg gggtcgacga  
27001 atcgcggcgc tgtccccgca ggaacagcgg gagacgttgc tgaccctcgt cggcgagacg  
27061 gtcgcggagg tgctgggaca cgagaccggc accgagatca acaccgctcg ggccttcagc  
27121 gaactcggcc tcgactcgct gggctcgatg gccctgcgtc agcgcctggc gggccgtacc  
27181 ggcctgcgga tgccggcctc gctggtcttc gaccaccgga cggtcaccgc gctcgcgcg  
27241 tacctgcgtc gactggctcg cggggactcc gacccgacc cggtacgggt gttcggcccc  
27301 accgacgagg ccgaaccctg gccctggttc ggcacccgct gccggttccc cggcgccatc  
27361 gccacccccg aggacctctg gcgggtgggtg tccgagggca cctccatcac caccggattc  
27421 cccaccgacc ggggctggga cctccggcg ctctaccacc ccgaccgga ccccccggc  
27481 accagctacg tcgacagggg gggattcctc gacggggccc cggacttcga ccccggttc  
27541 ttcgggatca cccccgcga ggcgctggcg atggaccgc agcagcggt caccctggag  
27601 atcgcggtgg aggcgggtgga acgggcgggc atcgaccgg agaccctcct cggcagcgac  
27661 accggcgtct tcgtcggcat gaacggccag tcctacctgc aactgctgac cggggagggt  
27721 gaccggctca accggtacca ggggttgggc aactcggcga gcgtgctctc cggcgtgtc  
27781 gcctacacct tcgggtggga ggggcggcg ctgacgggtg acaccgctg ctcgtcctcg  
27841 ctggtcgcca tccacctcgc catgcagtcg ctgctcggg gtgagtgctc gctggcggtg  
27901 gccggcgggg tgacgggtcat ggccgaccgg tacaccttcg tggacttcag cgcacagcgg  
27961 gggctcgccg ccgacggggc gtgcaaggcg ttctccgcg aggcgacgg gttcggcctc  
28021 gccgagggcg tcgcggcgct cgtcctcgaa ccgttgtcca aggcgcgcg aaacggccac  
28081 cagggtgctg cgggtgctgc cggcagcgcc gtcaaccagg acggggccag caacggcctc  
28141 gccgccccga accgggcgtc gcaggaacgg gtgatcaggc aggccttgac cgcctccggg  
28201 ctgctgcccg ccgacgtcga catgggtggag gcgcacggga cgggcaccga actcggcgac  
28261 ccgacgaggg ccggggcgct catcgcgcg tacggccggg accgggaccg gccgctctgg  
28321 ctgggctcgg tgaagacgaa catcgccac acccaggccg ccgccgggtg cgcgggggtg  
28381 atcaaggcgg tcctggcgat gcggcacggc gtactcccga ggtcgctgca cgcgacgag  
28441 ttgtccccgc acatcgactg ggcggacggg aaggctcgag tgctccgca ggcacgacg  
28501 tggccccccg gtgagcgccc ccgcccgcgc ggggtgtcct ccttcggcgt cagcgggacc  
28561 aacgcccacg tcacgtcga ggaggaccc gccgaaccgg accccgaacc ggttcccgc  
28621 gccccggggc ggcccctgcc cttcgtcctg cacggacgca gcgtccagac ggtccgggtc  
28681 caggcgcgga ccctcgccga acacctgcgc accaccggc accgggacct cgcgacacc  
28741 gcccgtaacc tggccaccgg tcgcgcccgt ttcgacgtcc gggccgcagt gctcggcacc  
28801 gaccgggagg gtgtctgcgc cgcctcgcac gcgctggcgc aggatcgccc ctcgcccagc  
28861 gtcgtcgccc cggcggtctt cgcgcgccgt acccccgctc tgggtcttccc cgggcagggg  
28921 tcgcagtggg tcggcatggc ccgtgacctg ctcgactcct ccgagggtgt cgcgagtcg  
28981 atgggcccgt gcgcgaggc gctgtcgccg tacaccgact gggacctgct cgacgtggtc  
29041 cgtgggggtc gcgacccccga cccgtacgac cgggtggacg tgctccagcc ggtgctgttc  
29101 gcgggtgatg tgctgctggc gcgggtgtgg cagtcgtacg ggggtgactc ggggtcggtg  
29161 gtgggtcact cgcaggggga gatcgccgc gcgcacgtgg ctgggtgcgt gtcgttggcc  
29221 gacgccgcca ggggtggggc gttgcgcagc cgggtgctgc gggagctcga cgaccagggc  
29281 ggcattgggt cggtcggcac ctcgccgcgc gaggttggact cggctcctgc ccgggtgggac  
29341 gggcggggtc cggtggcggc ggtgaacgga cccggcacgc tcgtgggtgg cggaccacc  
29401 gccgaactgg acgagttcct cgcgggtggc gaggcccgcg agatgaggcc gcgtcggatc  
29461 gcgggtgcgt acgcgtcgca ctccccggag gtggcccggg tcgaacagcg gctcggccgc  
29521 gaactcggca ccgtcaccgc cgtcggcggc acgggtcccgc tctactccac cgccaccggg  
29581 gacctcctcg acaccacagc catggacgcc ggggtactgt accgcaacct gcgccaaccg  
29641 gtgctgttcg agcacggcgt ccgcagcctc ctggagcggg gattcgagac gttcatcgag  
29701 gtcagccccg accctgtgct gctgatggcg gtcgaggaga ccgcgagga cgcgagcgc  
29761 ccggtcaccg gcgtgcccgc gctgcgcgc gaccacgacg ggccgtcggg gttcctccgc

29821 aacctcctgg gggcgacagt gcacgggggtc gacgtcgacc tgcgtccggc ggtcgcccac  
29881 ggccgcctgg tcgacctgcc cacctacccc ttcgacaggc agcggctctg gcccagccg  
29941 caccgcaggg ccgacacctc gtcgctgggg gtccgtgact cgaccacccc gctgctgcac  
30001 gccgcagtcg acgtacccgg tcacggcgga gcggtgttca ccgggcggct ctccccgac  
30061 gagcagcagt ggctgaccca gcacgtgggtg ggtgggcgga acctggtgcc cggcagtgte  
30121 ctggctcgacc tcgcgctcac cgccggggcc gacgtcggcg tgccgggtgct ggaggaaactc  
30181 gtcctgcagc agccgctgggt gttgaccgcc gccggtgctg tgctgcgcct gtcggtcggc  
30241 gccgccgacg aggacggggc gcgggccggtc gagatccacg ccgccgagga cgtctccgac  
30301 ccggccgagg ccgggtggtc ggcgtacgcg accgggaccc tcgccgtcgg cgtggccggc  
30361 ggccggccggg acggcacaca gtggcccccg ccgggcgcca ccgccctgac gttgaccgac  
30421 cactacgaca ccctcgccga actgggctac gagtacgggc cggcgttcca ggcgctgcgc  
30481 gccgcgtggc agcacggcga cgtgggtctac gcgaggtgt ccctcgacgc cgtcgaggag  
30541 ggggtacgct tcgacccgggt gctgctcgac gccgtcgccc agaccttcgg cctgaccagt  
30601 cgcgcccccg ggaagctccc cttegcctgg cggggcgta ccctgcacgc caccggggcc  
30661 actgcggtac ggggtgggtgg gacccccgcc ggaccggacg cgggtggccct gcgggtcacc  
30721 gaccgcacgg gtcagctcgt cgccacgggtg gacgcctgg tcgtcaggga cgccggggcg  
30781 gatcgggacc agccgcgcgg ccgcgacggc gacctgcacc gcctggagt ggtacggctg  
30841 gccaccccg acccgacccc ggccggcggtg gtgcacgtgg cggccgacgg gctcgacgac  
30901 ctgctgcgcg ccgggtgggtcc ggaccacag gccgtcgctg tccgctaccg tcccagcgcc  
30961 gacgacccga cggccgaggc ccgtcacggg gtgctctggg cggccacgct cgtgcgccgt  
31021 tggctcgacg acgaccgggtg gcccgccacc acctgggtgg tggccacgtc cgcaggggte  
31081 gaggtctccc ccggggagca cgtgccgcgc ccgggggccc ccgccgtgtg gggggtgctg  
31141 cgtgcgcgcc aggcggagtc ccgggacgc ttcgtgctcg tcgacggcga cccggagacg  
31201 ccccgggcgg tgccggacaa tccgcagctc gcggtccgtg acgggtgcgg gttcgtgcca  
31261 cggctgacgc cgctcgccgg tcccggtgcc gccgtcgccc accgggctga ccggctgggtg  
31321 cccggcaacg gcggctccat cgaggcagtg gccttcgccc ccgtccccga cgccgaccgg  
31381 cccctggcgc cggaggaggt acgcgtcgcc gtccgcgcca ccggcgtgaa ctccgtgac  
31441 gtcctgctcg cgctcggcat gtacccggaa ccggccgaga tgggcaccga ggcgtccggt  
31501 gtggtcaccg aggtcgggtc ggggtgtccg cggttcacc ccggccaggc ggtgacgggc  
31561 ctgttccagg gggccttcgg gccgggtggcg gtcgccgacc accggctcct caccgccgtc  
31621 cccgacgggt ggccggcggt ggacgccgca gccgtaccca tcgcgttcac caccgcccac  
31681 tacgcgctgc acgacctggc cgggttgtag gccgggcagt ccgtgctgggt ccacgcgcgc  
31741 gccggcgggg tggggatggc tgccgtcgcg ttggcccgct gggccggggc ggagggtgtt  
31801 gccacggcca gcccgggcaa acaccgacg ctgcggggcg tcggcctcga cgacgaccac  
31861 atgcctcgt ccggggagag cgggttcggt gagcggttcg ccgcgcgtac cggggggcgg  
31921 ggctcgacg tggctcctgaa ctgcctcacc ggcgacctgc tcgacgagtc cgcgcggctg  
31981 ctgcgcgacg gcgggggtctt cgtcgagatg ggcaagaccg acctgcggcc ggccggagcag  
32041 ttccggggcc ggtacgtccc gttcgacctg gccgaggccg gtcccgatcg gctcggcgag  
32101 atcctggagg aggtcgtcgg tctgctggcc gccggtgccc tcgaccgggt gcccgtgtcg  
32161 gtgtgggagt tgtcggcggc cccggcccgc ctcaccacaa tgagccgggg ccgacacgtg  
32221 ggcaagctcg tcctcaccca gcccgccccc gtgcaccccc aggaaacgg gctggtcacc  
32281 ggccgggacc gcaccctggg gcggctggtc gcccgccacc tggtgaccgg gcacggcgta  
32341 cccacacctc tggtgggcag ccggcgcggt ccggcgggcc cgggcgcggc cgagctgcgc  
32401 gccgacgtcg aaggcctcgg cgcgaccatc gagatcgctg cctgcgacac cgccgaccgg  
32461 gaggcgctcg cggcgctgct cgactcgatc ccgcgggacc gtccgctgac cggggtgggtg  
32521 cacaccgccc gggctcctggc cgacgggctg gtcacctcca tcgacgggac cgccaccgat  
32581 caggctcctgc gggccaaggt cgacgcggcg tggcacctgc acgacctgac ccgggacgcg  
32641 gacctgagct tcttcgtgct gttctcgtcg gcggcgctcg tgctggccgg tcccgggcag  
32701 ggctgttacg cggcgggcaa cggggtcctc aacgccctgg ccgggcaacg gcggggccctc  
32761 ggactgcccg cgaaggcgct cgggtggggc ctgtggggcg aggccagcga gatgaccagc  
32821 ggctcgggtg accggatcgc ccgtaccggg gtcgccgcgc tgccgaccga gcgggcgctg  
32881 gccctgttcg acgcgggtct gcgcagcggc ggggaggtgc tgttcccgt gtctgtcgac  
32941 aggtcggcgc tgcgcggggc cgagtacgtc ccgaggtgc tgccggcgcc ggtccggctc  
33001 acgccacggg ccgccaacag ggccgagacc ccggggccgg gcctgctcga ccgtctcgtc  
33061 ggtgcacccg agaccgatca ggtggccgcg ctggccgagc tgggtccgct gcacgcggcg  
33121 gcggctcgcc gctacgactc ggccgaccag ctgcccgaac gcaaggcgtt caaggacctc  
33181 gggttcgact cgctggcggc ggtggagctg cgcaaccggc tcggcgctac caccggcgta  
33241 cggctgcccga gcacgctgggt gttcgaccac ccgacaccgc tggcggtggc cgaacacctg  
33301 cggctcggagt tgttcgcccga ctccgcgcgc gacgtcgggg tcggtgcgcg cctcgacgac  
33361 ctgggaacgg cgctcgacgc cctgcccgcg gcgcagggac acgccgacgt cggggcccgc  
33421 ctggaggcgc tgctgcgcgc gtggcagagc cgacgacccc cggagaccga gccagtgcag  
33481 atcagtgcag acgccagtga cgacgagctg ttctcgatgc tcgacaggcg tctcggcggg  
33541 ggagggggac tctaggtgac aggtcgatcc cgccccgcgg cagtggaccg taccgccctg  
33601 acaggctccac cgggttcgcg tcgcctccca caccgacgg ccggggtatc cacggaagg



33661 atccgatgag cgagagcagc ggcatgaccg aggaccgcct ccggcgetat ctcaagcgca  
33721 ccgtcgccga actcgactcg gtgacaggtc ggctcgacga ggtcgagtac cgggcccgcg  
33781 aaccgatcgc cgtcgtcggc atggcctgcc gggtcccccg ggggtgtggac tcgcccggagg  
33841 cgttctggga gttcatccgc gacgggtggtg acgcgatcgc cgaggcgccc acggaccgtg  
33901 gctggccgcc ggcaccgcga cccgcctcgt gtgggtctcct cgcggagccg ggcgcgttcg  
33961 acgcccctt ctteggcatc tcaccccgcg aggcgctcgc gacggacccc cagcagcgcc  
34021 tgatgctgga gatctcctgg gaggcgttgg agcgtgcggg tttegacccg tcgagcctgc  
34081 gcggcagcgc cgggtggcgtc ttcaccgggtg tcgggtgcggg ggactacgga cccaggcccg  
34141 acgaggcacc cgaggagggtg ctccggctacg tcggcatcgg caccgcctcc agcgtcgcc  
34201 ccggacgggt ggcgtacacc ctgggggttgg aggggtccagc cgtcaccgtc gacaccgcct  
34261 gctcctccgg gctcaccgcg gtgcacctgg cgatggagtc gctgcgcgcg gacgagtga  
34321 ccctggctct cgcgggtggg gtcaccgtga tgagcagccc ggggtgcgttc accgagttcc  
34381 gcagccaggg cgggttggcc gaggacggcc gctgcaaacc gttctccgcg gccgcgcagc  
34441 gcttcgggct cgcggagggg gccgggggtcc tgggtgctcca acggctgtcc gtcgcccggg  
34501 ccgagggccg gccgggtgctg gccgtactgc gtgggtcggc gatcaaccag gacgggtgca  
34561 gcaacgggct caccgcgcgc agcggccccc cccagcggcg ggtgatcagg caggcgttgg  
34621 agcgggcgcg gctgcgtccc gtcgacgtgg actacgtgga ggcccacggc accggcaccc  
34681 ggctggggcg tccgatcgag gcgcacgccc tgctcgacac gtacgggtgc gaccgggaac  
34741 ccggccgccc gctctgggtc ggatcggtga agtccaacat cggtcacacc caggcggcgg  
34801 cgggggtggc cgggggtgatg aagaccgtgc tggcgtgcg gcacgggag atcccggcga  
34861 cgttgcactt cgacgagccc tcgcccgcag tcgactggga ccgggggtgc gtgtcggtgg  
34921 tgtccgagac ccggccctgg ccgggtgggg agcgcgcgcg ccggggcggg gtgtcctcgt  
34981 tcggcatcag ccggaccaac gcgcacgtca tcgtcgagga ggccgcgagc ccgcaggcgg  
35041 ccgacctcga cccgaccccc ggcccggcaa ccggagcgac ccccggaaac gatgccgccc  
35101 ccaccgccga gccgggtgcg gaggcggctc cactggtgtt ctccgcgcgc gacgagcggg  
35161 ccctgcgcgc ccaggcggcc cggctcgcgc accgtctcac cgacgacccg gcccctcgt  
35221 tgcgcgacac cgccttcacc ctgggtcacc gccgtgccac ctgggagcat cgggcggctc  
35281 tcgtcggcgg gggcgaggag gtccctcgcg gcctccgggc cgtcgccggg ggacgtccc  
35341 tcgacggagc cgtcagcggg cgggcgcgcg ccggccgcgc ggtgggtgct gtcttccc  
35401 ggcaggggcg acagtggcag ggcatggccc gggacctgct gcggcagtc cccgacctc  
35461 cggagtccat cgacgcctgc gagcgggcgc tcgccccgca cgtggactgg tcgctgcgc  
35521 aggtgctcga cggcgagcag tcgttggacc ccgtcgacgt ggtgcagccg gtgtgttc  
35581 cggtgatggt gtcgttggcg cgggtgtggc agtcgtacgg ggtgactccg ggtgcgggtg  
35641 tgggtcactc gcagggggag atcgccgcgc cgcacgtggc tgggtgcgtt tcgttggccg  
35701 acgcccagc ggtgggtggc ttgcgcagcc ggggtgctgc ccgtctcggg ggtcacggcg  
35761 ggatggcgtc gttcggggtc caccgccgac aggcgcgcga gcggatcgc cgttcgcgcg  
35821 gtgcgctgac tgcgcctcgc gtcaacggtc cccgttcggg ggtgctggcc ggggagaacg  
35881 gcccggtgga cgagctgatc gccgagtgcg aggcgcgagg cgtgaccgcc cgtcggatcc  
35941 ccgtcgacta cgcctcacac tcccgcagc tggagtgcgt gcgtgaggag ctgctcgccg  
36001 cactggcccg ggtccgtccg gtgtcggccg ggatccccct gtactcgacc ctgaccggtc  
36061 aggtcatcga aacggcgacg atggacgcgc actactggtt cgccaacctc cgggagccgg  
36121 tgcgcttcca ggacgccacc aggcagctcg ccgaggcggg gttcgacgcc ttcgtcgagg  
36181 tcagcccga cccgggtgtg acagtccgtg tcgaggccac cctcgaggca gtgtgcgcc  
36241 ccgacgcgga tccgtgtgtc acaggcaccc tgcgcgcga acgcggcggt ctgcgcagc  
36301 tccacaccgc gctcgccgag gcgtacaccc ggggggtgga ggtcgactgg cgtaccgcag  
36361 tgggtgaggg acgcccggtc gacctgcgg tctaccggt ccaacgacag aacttctggc  
36421 tcccggctcc cctggggccg gtccccgaca ccggcgacga gtggcggtac cagctcgcc  
36481 ggcaccccgt cgacctcggg cggctcctccc tggccggacg ggtcctgggt gtgaccggag  
36541 cggcagtacc cccggcctgg acggacgtgg tccgcgacgg cctggaacag cgcggggcga  
36601 ccgtcgtgtt gtgcaccgcg cagtcgcgcg ccggatcgg cgcgcgactc gacgccgtc  
36661 acggcaccgc cctgtccact gtggtctctc tgctcgcgct cgcggagggc ggtgctgtc  
36721 acgaccccag cctggacacc ctgcggttgg tccaggcgt cggcgagcc gggatcgacg  
36781 tcccctgtg gctgggtgac agggacgcgc ccgcgtgac cgtcgagagac gacgtcgatc  
36841 cggcccaggc catgggtcgt gggctcggcc ggggtgggtgg cgtggagtcc cccgcccgg  
36901 ggggtggcct ggtggacctg cgcgaggcc agcccgactc ggcccggctc ctggccgcca  
36961 tactggccga cccgcgcggc gaggagcagt tcgcgatccg gcccgacggc gtcaccgtc  
37021 cccgtctcgt cccggcaccc gccgcgcgcg cgggtacccg gtggacgcgc cgcgggaccg  
37081 tcctggctac cggcggcacc ggcggcatcg gcgcgcacct ggcccgtggt ctcgcccgtg  
37141 cgggcgcgca gcacctggtg ctgctcaaca ggccggggagc ggaggcggcc ggtgcgcgcg  
37201 acctgcgtga cgaactggtc gcgctcggca cgggagtcac catcacggcc tgcgacgtc  
37261 ccgaccgcga ccgggttggcg gccgtcctcg acgcgcacg ggcgcaggga cgggtggtca  
37321 cggcggtgtt ccacgcgcgc gggatctccc ggtccacagc ggtacaggag ctgaccgaga  
37381 gcgagttcac cgagatcacc gacgcgaagg tgcgggggtac ggcgaacctg gccgaactc  
37441 gtcccagact ggacgcctc gtgctgttct cctcgaacgc ggcgggtgtg ggcagcccg

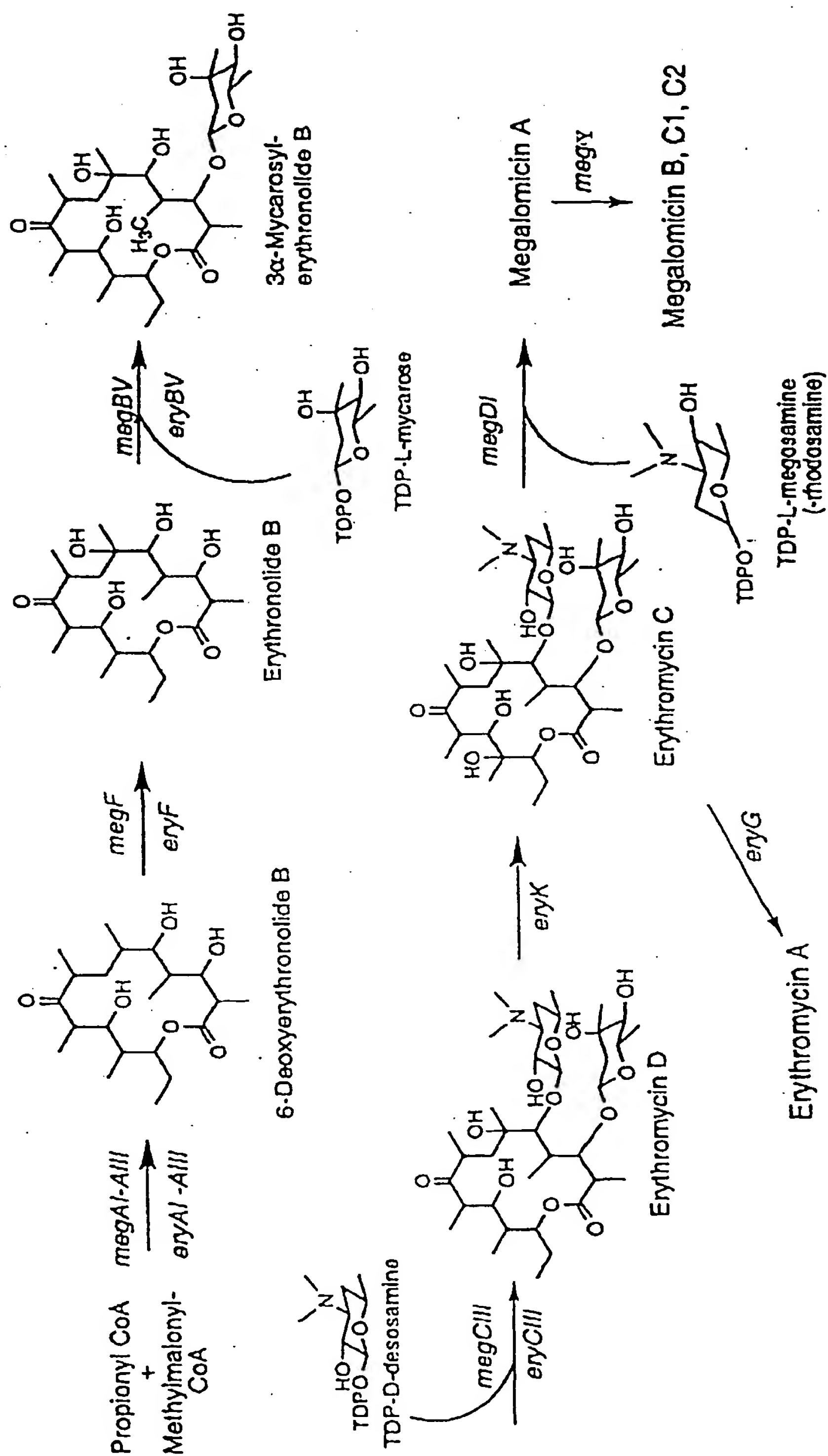
37501 ggctggcctc ctacgcggcg ggcaacgcct tcctcgacgc cttcgcccgt cgtggtcggc  
37561 gcagtgggct gccggtcacc tcgatacgct ggggtctgtg ggccgggag aacatggccg  
37621 gtaccgaggg cggcgactac ctgctgcagcc agggcctgcg cgccatggac ccgcagcggg  
37681 cgatcgagga gctgcggacc accctggacg ccggggaccc gtgggtgtcg gtgggtggacc  
37741 tggaccggga gcgggttcgtc gaactgttca ccgccgccc cgccggccc ctcttcgacg  
37801 aactcgggtg ggtccgcgcc ggggcccagg agaccgggtca ggaatcggtat ctgcccggc  
37861 ggctggcgct gatgccggag gccgaacgtc acgagcatgt cgcccggtg gtccgagccg  
37921 aggtggcagc ggtgctgggc cacggcacgc cgacgggtgat cgagcgtgac gtccgcttc  
37981 gtgacctggg attcgactcc atgaccgccc tcgacctgcg gaaccggctc gcggcgggtga  
38041 ccgggggtccg ggtggccacg accatcgtct tcgaccaccc gacagtggac cgcctcaccg  
38101 cgcactacct ggaacgactc gtcgggtgagc cggaggcgac gaccccggtc gcggcgggtc  
38161 tcccgcaggg acccggggag gccgacgagc cgatcgcgat cgtcgggatg gcctgccgc  
38221 tcgccgggtg agtgcgtaacc cccgaccagt tgtgggactt catcgtcgcc gacggcgacg  
38281 cgggtcaccga gatgccgtcg gaccgggtcct gggacctcga cgcgctgttc gacccggacc  
38341 ccgagcggca cggcaccagc tactcccgcc acggcgcggt cctggacggg gcggccgact  
38401 tcgacgcggc gttcttcggg atctcgccgc gtgaggcggt ggcgatggat ccgcagcagc  
38461 ggcaggtcct ggagacgacg tgggagctgt tcgagaacgc cggcatcgac ccgcactccc  
38521 tgcgcgggtac ggacaccggt gtcttcctcg gcgctgcgta ccagggggtac ggccagaacg  
38581 cgcaggtgcc gaaggagagt gagggttacc tgctcaccgg tggttcctcg gcggtcgct  
38641 ccggtcggat cgcgtacgtg ttgggggttg aggggcccgc gatcactgtg gacacggcgt  
38701 gttcgctcgtc gcttgtggcg ttgcacgtgg cggccgggtc gctgcgatcg ggtgactgtg  
38761 ggctcgcggg ggccgggtgg gtgtcgggtga tggccgggtc ggagggtgtc accgagttct  
38821 ccaggcaggg cgcgctggcc cccgacgggt ggtgcaagcc cttctccgac caggccgacg  
38881 ggttcggatt cgccgagggc gtcgctgtgg tgctcctgca gcgggtgtcg gtggcgggtg  
38941 gggagggggc tcgggtgttg ggtgtgggtg tgggttcggc ggtgaatcag gatggggcga  
39001 gtaatgggtt ggcggcgccg tcgggggttg cgcagcagcg ggtgattcgg cgggcgtggg  
39061 gtcgtgcggg tgtgtcgggt ggggatgtgg gtgtgggtga ggcgcattgg acggggacgc  
39121 ggttggggga tccgggtggg ttgggggctg tgttggggac gtatgggggt ggtcgggggt  
39181 ggggtgggtc ggtgggtggg ggttcgggtga aggcgaatgt ggggtcatgt caggcggcgg  
39241 cgggtgtggg ggggtgtgat aagggtgggt tgggggttg gggtgggttg gtgggtccga  
39301 tgggtgtgtc ggggtgggtt tcgggggttg tggattgggt gcggtgggtg ttgggtgggt  
39361 cggatggggg gcgggggttg ccgggtgggt tggatgggg gcgtcggggg ggggtgtcgg  
39421 cgtttggggg gtcggggacg aatgctcatg tggtgggtgg ggaggcggc gggtcgggtg  
39481 tggggggcga acggccgggt gaggggtcgt cgcgggggtt ggtgggggtg gctgggtggg  
39541 tgggtgccgg ggtgctgtcg gcaaagaccg aaaccgccc gaccgagctc gcccgacgac  
39601 tgcacgacgc cgtcgacgac accgtcgccc tcccgggcgg tgcccgccacc ctgcccaccg  
39661 gacgcgccc cctgcccctac cgggcccggc tgctggccc cgaccacgac gaactgcgcg  
39721 acaggctgcg ggcgttcacc actgggttcgg cggtcccgg tgtgggtgtc ggggtggcgt  
39781 cgggtgggtg tgtgggtgtt gtttttctct gtcagggtgg tcagtgggtg gggatggcgc  
39841 ggggggttgt gtcgggttcg gtgtttgtgg agtcgggtgg ggagtgtgat gcgggtgggt  
39901 cgtcgggtgg ggggttttcg gtgttggggg tgttggaggg tcggtcgggt gcgccgtcgt  
39961 tggatcgggt ggatgtgggt cagccgggtg tgttcgtggg gatgggtgtc ttggcgcggg  
40021 tgtggcgggt gtgtgggggt gtgcctgcgg cgggtgggtg tcatcgcag ggggagatcg  
40081 cggcggcggg ggtggcgggg gtgttgtcgg tgggtgatgg tgcgcgggtg gtggcgttgc  
40141 gggcgcgggc gttgcggggc ttggccggcc acggcggcat ggtctccctc gcgggtctcc  
40201 ccgaacgcgc ccgggagctg atcgcaccct ggtccgaccg gatctcgggt cctcgccgcc  
40261 actccccgac ctcggtgggt gtctcgggtg acccacaggc ggactacgcc tcccactccg  
40321 actgcgccga gaccgggtgag cgggccaaga cgctgcctgt ggccgacgtc acggcgcgcc  
40381 cccacgtcga acagatccgc gacacgatcc tcaccgacct gggcgccggc acggacatgg  
40441 gacccgacgt cgccctctac tccacgctgc acggcgccc caccgggtcg cctcgacgag  
40501 acgcccggta ctggtacgac aacctgcgtc tcgagatgag cccacacccg ctcgctgcac  
40561 ccgcccgtcg cgacgggtac cgggtcttcg tggccatcgg gggcccacgt gcacggcgta  
40621 ccgcccgtgca ggagatcgac gacgagacgg tggccacgtt gaactacccg ttcgaggcga  
40681 gcgagcggca cctggtcgcc gaactcgccc ttcccctgcc cagaccacgc taccgcgtcg  
40741 ggcggggcgt cctccccgcc acccaccggg agctgtccgg cgggctcctc gtcccgggtg  
40801 cccgggtact gctcgccccg acggcggcgg agctgtccgg cgggctcctc gtcccgggtg  
40861 actggcggcc cctggccacc accccggcgg cacagcgtcg cctggacga ggcggcggga  
40921 acgccccgga gaccctcggc ctcgcggtcg cccacctggc cgggacccga ctcctcggcg  
40981 ccgctcccga ccgggagtcg gacaccgcca ctctgggtgg tcaccagcgg cggacgggtg  
41041 gtgtgctctc cttcgccgcc caggcaatgg tgtgggggat atgggtctgg  
41101 agggcgacgt cgaggcccca cggctgggtg acgtgaccgt cgaaccaccc gccgaggacg  
41161 acgacccgat cgactgcgac ctggccgccc acgaccacga ggcctgcgcg  
41221 agaccccgca ccgggtgggg ctggccgccc acgaccacga ggcctgcgcg  
41281 ggggtgggtct cgcgcgccc

41341 acggcatccg ccacggccga cggctcgtec gcgccccgct gaccaccga aacgccaggt  
41401 ggacaccggc gggcacggcg ctcgtcacgg gcggtacggg tgccctcggc ggccacgtcg  
41461 cgcggtacct ggcccgggtc ggggtgaccg atctcgctct gctcagcagg agcggccccg  
41521 acgcacccgg tgccgccgaa ctggccgccc aactggccga cctcggggcc gagccgagag  
41581 tcgaggcggt cgacgtcacc gacggggccac gcctgcgcgc cctggtgcag gagctacggg  
41641 aacaggaccg gccgggtccg atcgctcgtec acaccgcagg ggtgcccgaac tcccgteccc  
41701 tcgaccggat cgacgaactg gagtcgggtca gcgcccgcga ggtgaccggg gcgcggtgc  
41761 tcgacgagct ctgcccggac gccgacacct tcgctcctgtt ctccctcggg gcgggagtgt  
41821 ggggtagcgc gaacctgggc gcgtacgcgg cagccaacgc ctacctggac gccctggccc  
41881 accgccggcg ccaggcgggc cgggcccgcga cctcgggtcg ctggggggcg tgggccggcg  
41941 acggcatggc caccggcgac ctcgacgggc tgaccggcg cggctcgcgg gcgatggcac  
42001 cggaccgggc gctgcgcgcc tgcaccaggc gttggaccac ccacgacacc tgtgtgtcgg  
42061 tagccgacgt cgactgggac cgcttcgccc tgggtttcac cgccgcccg cccagacccc  
42121 tgategacga actcgtcacc tccgcgcggc tggccgcccc caccgctgcg gcggccccgg  
42181 tcccggcgat gaccggcgac cagctactcc agttcacgcg ctcgcacgtg gccgcgatcc  
42241 tcggtcacca ggacccggac gcggtcgggt tggaccagcc ctccaccgag ctgggcttcg  
42301 actcgctcac cgccgtcggc ctgcgcaacc agctccagca ggccaccggg cggacgtgc  
42361 ccgccgccct ggtgttccag caccacacgg tacgcagact cgccgaccac ctgcgcgacg  
42421 agctcgacgt cggcaccggc ccggtcggag cgacgggcag cgtcctgcgg gacggctacc  
42481 ggcggggcgg gcagaccggc gacgtccggt cgtacctgga cctgctggcg aacctgtcgg  
42541 agttccggga gcggttcacc gacgcggcga gcctggggcg acagctggaa ctcgtcgacc  
42601 tggccgacgg atccggcccc gtcactgtga tctgttgccg gggcactgcg gcgctctccg  
42661 ggccgcacga gttcgccccg ctgcctcgg cgctgcgcgg caccgtgccg gtgcgcgcc  
42721 tcgcgcaacc cgggtacgag gcgggtgaac cggtgccggc gtcgatggag gcagtgtcgg  
42781 ggggtgcaggc ggacgcggtc ctgcggcac agggcgacac gccgttcgtg ctggtcggac  
42841 actcggcggg ggccctgatg gcgtacgccc tggcgaccga gctggccgac cggggccacc  
42901 cgccacgtgg cgtcgtgctc ctgcacgtgt acccaccgg tcaccaggag gcggtgcacg  
42961 cctggctcgg cgagctgacc gccgcctgt tcgaccacga gaccgtacgg atggacgaca  
43021 cccggctcac ggccctgggg gcgtacgaca ggctgaccgg caggtggcgt ccgagggaca  
43081 ccggtctgcc cacgtgggtg gtggccgcca gcgagccgat gggggagtgg ccggacgacg  
43141 gttggcagtc cacgtggccg ttcgggcacg acagggtcac ggtgcccggg gaccacttct  
43201 cgatggtgca ggagcacgcc gacgcgatcg cgcgccacat cgacgcctgg ttgagcgggg  
43261 agagggcatt aacacgaccg atcgcgccgt gctggggcca cgactccaga tgatccgggg  
43321 actgtactgg ggttacggca gcaacggaga cccgtaccgg atgctgttgt gcgggcacga  
43381 cgacgacccg caccgctggt accggggggt gggcggtacc ggggtccggc gcagccgtac  
43441 cgagacgtgg gtggtgaccg accacgccac cgccgtgcgg gtgctcgacg acccgacctt  
43501 caccggggcc accggccgga cgccggagtg gatgcggggc gcgggcgccc cggcctcgac  
43561 ctgggcccag ccgttcctgt acgtgcacgc cgctcctgg gacgcggaac tgcccgaacc  
43621 gcaggaggtg gaggaccggc tgacgggtct cctgcctgcc ccgggggacc gcctggacct  
43681 ggtccgcgac ctgcctggc cgatggcgct gcgggggggt ggcgcgacg accccgacgt  
43741 gctgcgcgcc gcgtgggacg cccgggtcgg cctcgacgcc cagctcacc cgagccctt  
43801 ggcgggtgacc gaggcggcga tcgcccgggt gcccggggac ccgcaccggc gggcgctgtt  
43861 caccgccgtc gagatgacag ccaccgcgtt cgtcgacgcg gtgctggcg tgaccgccac  
43921 ggcggggggc gccagcgctc tcgcccgaac ccccgacgtc gccgcccgtc tcgtcgcgga  
43981 ggtgctgcgc ctgcatccga cggcgcacct ggaacggcgt accgcccggc ccgagacggt  
44041 ggtgggcgag cacacggtcg cggcgggcga cgaggtcgct gtggtggtcg ccgcccga  
44101 ccgtgacgcg ggggtcttcg ccgaccggga ccgcctcgac ccggaccggg ccgacgccga  
44161 ccgggcccctg tccgcccgac gcggtcacc cggccgggtg gaggagctgg tgggtgtcct  
44221 gaccaccgcc gactgcgca gcgtcgccaa ggcgtgcc ccggtgccc ccggtggccc  
44281 ggtcgtcagg cgacgtcggt caccggctct gcgagccacc gccactgcc cggtcgaact  
44341 ctgaggtgcc tgcgatgcgc gtcgtcttct cctccatggc cagcaagagc cacctgttcg  
44401 gtctcgttcc cctgcctgg gccttcgcg cgggcgggcca cgaggtacgg gtcgtcgctt  
44461 caccggctct caccgacgac atcacggcgg ccggactgac ggccgtaccg gtcggcaccg  
44521 acgtcgacct tgtcgacttc atgaccacg ccgggtacga catcatcgac tacgtccgca  
44581 gcctggactt cagcgagcgg gacccggcca cctccacctg ggaccacctg ctcggcattg  
44641 agaccgtcct caccgccgac ttctacgccc tgatgagccc ggactcgctg gtcgagggca  
44701 tgatctcctt ctgtcggtcg tggcgaccgg actggctgct tggaccgcag accttcgccc  
44761 cgtcgatcgc ggcgacgggt accggcggtg cccacgcccc actcctgtgg ggaccgcaca  
44821 tcacggtacg ggcccggcag aagttcctcg ggctgctgcc cggacagccc gccgcccacc  
44881 gggaggaccc cctcgccgag tggctcacct ggtctgtgga gaggttcggc ggccgggtgc  
44941 cgcaggacgt cgaggagctg gtggtcgggc agtggacgat cgaccccgcc ccggtcggga  
45001 tgcgcctcga caccgggctg aggacgggtg gcatgcgcta cgtcgactac aacggcccgt  
45061 cgggtggtgcc ggactggctg caccgacgag cgaccggcg acgggtctgc ctaccctgg  
45121 gcatctccag ccgggagaac agcatcgggc aggtctccgt cgacgacctg ttgggtgcgc

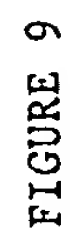


45181 tcggtgacgt cgacgccgag atcatcgca cagtggacga gcagcagctc gaaggcgtcg  
45241 cccacgtccc ggccaacatc cgtacggctc gggttcgtccc gatgcacgca ctgctgccga  
45301 cctgcgcggc gacgggtgcac cacggcggtc ccggcagctg gcacaccgcc gccatccacg  
45361 gcgtgccgca ggtgatcctg cccgacggct gggacaccgg ggtccgcgcc cagcggaccg  
45421 aggaccaggg ggcgggcatc gccctgccgg tgcccagact gacctccgac cagctccgcg  
45481 aggcgggtgcg gcgggtcctg gacgatcccg ccttcaccgc cgggtgcggcg cggatgcggg  
45541 ccgacatgct cgccgagccg tccccgcgcg aggtcgtcga cgtctgtgcg gggctggctg  
45601 gggaacggac cgccgtcgga tgagcaccga cgccaccac gtccggctcg gccgggtgcg  
45661 cctgctgacc agccggctct ggctgggtac ggcagccctc gccggccagg acgacgccga  
45721 cgcagtacgc ctgctcgacc acgcccgttc ccggggcgctc aactgcctcg acaccgccga  
45781 cgacgactct gcgtcgacca gtgcccaggt cgccgaggag tcggtcggcc ggtggttggc  
45841 cggggacacc ggtcggcggg aggagaccgt cctgtcgggt acggtgggtg tcccaccggg  
45901 cgggcaggctc ggcggggggcg gcctctccgc ccggcagatc atgcctcct gtgagggtc  
45961 cctgcggcgt ctgggtgtcg accacgtcga cgtccttcac ctgccccggg tggaccgggt  
46021 ggagccgtgg gacgaggtct ggacggcggt ggacgccctc gtggccgccg gaaaggtctg  
46081 ttacgtcggg tcgtcgggt tccccggatg gcacatcgte gccgccagg agcacgccgt  
46141 ccgccgtcac cgcctcggcc tgggtgccca ccagtgtcgg tacgacctga cgtcgcgcca  
46201 tcccgaactg gaggtcctgc ccgccgcgca ggcgtacggg ctcggggtct tcgccaggcc  
46261 gaccgcctc ggcggtctgc tcggcggcga cggtcggggt gccgcagccg cacgggcgtc  
46321 gggacagccg accggcactgc gctcggcggt ggaggcgtac gaggtgttct gcagagacct  
46381 cggcgagcac ccgccgagg tcgcaactggc gtgggtgtct tcccggcccc gtgtggcggg  
46441 ggcggtcgtc ggtgcgcgga cgcccggacg gctcgaactc gcgtcccgcg cctgcggcgt  
46501 cgccctcggc gcgacggaac tcaccgccct ggacgggate ttccccgggg tcgccgcagc  
46561 aggggcggcc ccggaggcgt ggctacgggt agagcccgc cctgacctgc gggaaaccgt  
46621 gtcgggtgcg cgggacggcc gccgcgggtc ccgcccggt cagccgggtg gggtagaccg  
46681 cagcaggtcc ggcgccaccg actcggccac ctccccgacg tggtcggcga ggtagaagtg  
46741 cccgcccggg aaggtccggg tacggccggg gactaccgag tacggcagcc agcgttgggc  
46801 gtcctccacc gtcgtcaacg ggtcgggtgt accgcagagg gtggtgatgc cggcccgcag  
46861 cggcggcccc gcctgccagg cgtaggagcg cagcaccggg tggtcggccc gcagcaccgg  
46921 cagcgacatg tccaacagcc cctggtcggc caatgcggcc tcgctgacct cgagcctgcg  
46981 catctgctcg acgagtcctg cctcgtcggg caggtcgggt cgccgctcgt ggaccggggg  
47041 ggcgggtctgc ccggagacga acaaccgcag cggtcgcacc cccggacgag cctccaggcg  
47101 acgggcgggtc tcgtaggcga ccagggcgcc catgctgtga ccgaacaggg cgaacggaac  
47161 ctgcccgacg aggtcgcgca gcacggccgc gacctcgtcg gcgatctccc cggcggtgcc  
47221 gagagcccg cgtcacgtc ggtcctgcgg gcccggtac tgcaccgcc acacgtcgac  
47281 ctccggggcc agtgcccggg cgaggctcag gtacgagtcg gcggcggtc ccgcgtgcgg  
47341 gaagcagtac agccgggccc ggtgtccgtc ggcggacccg aaccgccgca accaggtgtt  
47401 catcggtgtc tcatccgttc ggtcgcaccg gcaggtggtc gatgccgcgc agcaggagcg  
47461 accgccgcca gacaacctcg tcggagggga agcccagcga cagcttcggg aagcggtcga  
47521 acagggcccc cagggcgacc tctccctcca gcttggccag cgggcggccc atgcagtagt  
47581 ggatgccgtg cccgaagggt aggtgtcccc ggctgtccct ggtgacgtcg aaccggtcgg  
47641 ggtcggggaa ctgtcccggg tcggggttgg ccgcccggtt ggcgatcagg acggtgctgt  
47701 acgccgggat cgtcaccccg ccgatctcca cctcggcggt ggcgaaccgg gtggtggtct  
47761 ccggtggggc ctggtagcgc aggatctcct ccaccgctcc gggcagcagt gccgggtcct  
47821 tccggaccag cgcgagctgg tcggggtggg tcagcagcag gtaggtgccg atcccgatga  
47881 ggctcaccga cgcctcgaat cccgccagca gcagcaccag cgcgatggag gtgagttcgt  
47941 cgcggctgag ccggtcggcg tcgtcgtcct ggaccggat c

(SEQ ID NO: 1)



**FIGURE 8**





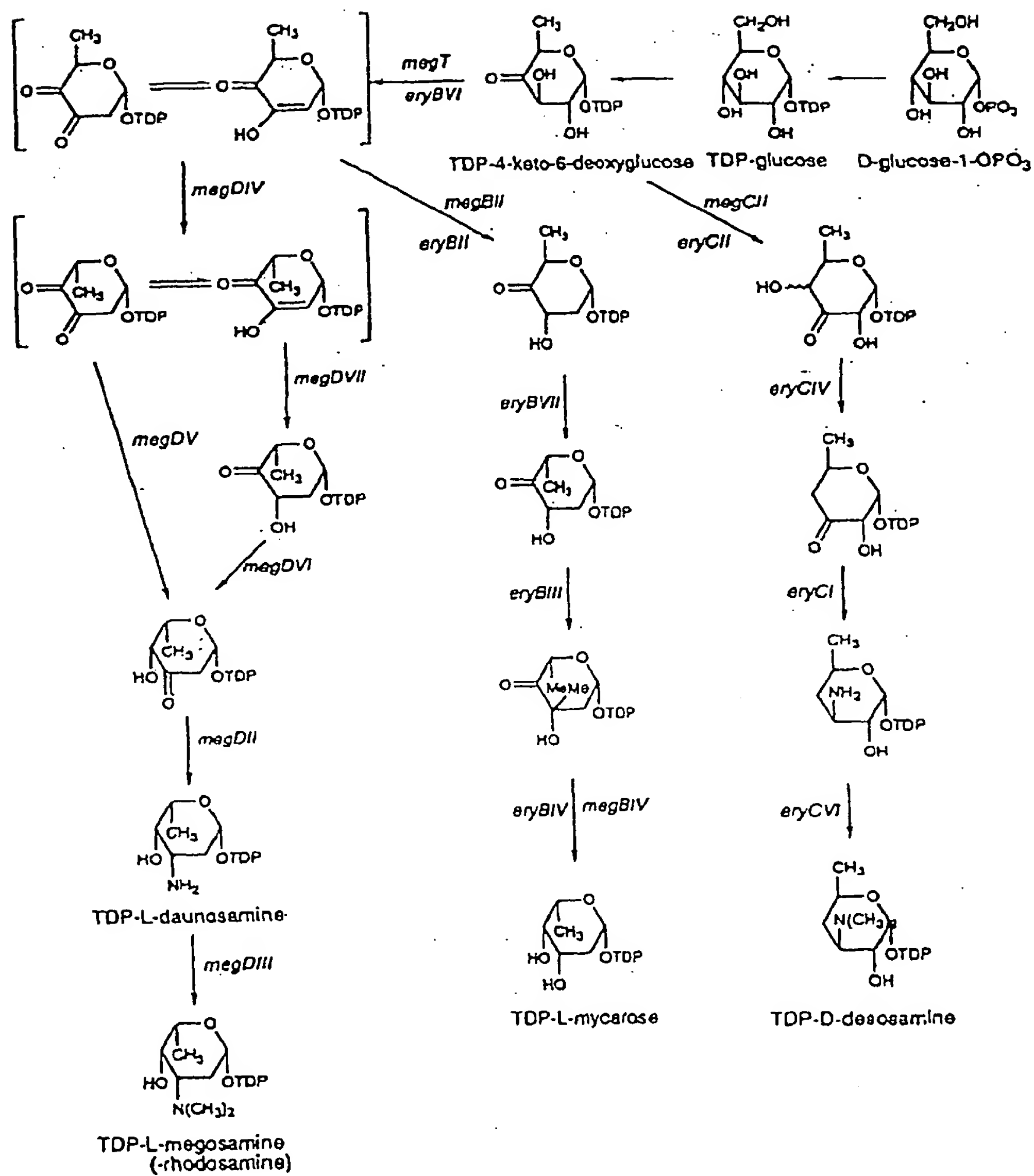


FIGURE 10

## SEQUENCE LISTING

<110> Kosan Biosciences, Inc.

<120> Recombinant Megalomycin Biosynthetic  
Genes and Uses Thereof

<130> 300622004740

<140> To be assigned

<141> Herewith

<150> US 60/158,305

<151> 1999-10-08

<150> US 60/190,024

<151> 2000-03-17

<160> 34

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 47981

<212> DNA

<213> Micromonospora megalomicea

<220>

<221> CDS

<222> (1)...(144)

<223> megBVI(megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
SEQ ID NO: 2= translated amino acid sequence

<221> CDS

<222> (928)...(2061)

<223> megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,  
TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
SEQ ID NO: 3= translated amino acid sequence

<221> CDS

<222> (2072)...(3382)

<223> megDI, rhodosaminyl transferase (eryCIII homolog),  
TDP-megosamine glycosyltransferase;  
SEQ ID NO: 4= translated amino acid sequence

<221> CDS

<222> (3462)...(4634)

<223> megG(megY), mycarosyl acyltransferase, mycarose O-acyltransferase;  
SEQ ID NO: 5= translated amino acid sequence

<221> CDS

<222> (4651)...(5775)

<223> megDII, deoxysugar transaminase (eryCI, DnrJ homolog),  
TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
SEQ ID NO: 6= translated amino acid sequence

<221> CDS

<222> (5822)...(6595)

<223> megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
SEQ ID NO: 7= translated amino acid sequence

<221> CDS  
<222> (6592)...(7197)  
<223> megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU homolog), TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
SEQ ID NO: 8= translated amino acid sequence

<221> CDS  
<222> (7220)...(8206)  
<223> megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),  
TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 9= translated amino acid sequence

<221> CDS  
<222> (8228)...(9220)  
<223> megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
SEQ ID NO: 10= translated amino acid sequence

<221> CDS  
<222> (9226)...(10479)  
<223> megBV, mycarosyl transferase, mycarose glycosyltransferase;  
SEQ ID NO: 11= translated amino acid sequence

<221> CDS  
<222> (10483)...(11424)  
<223> megBIV, TDP-hexose 4-keotoreductase,  
TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12= translated amino acid sequence

<221> CDS  
<222> (12181)...(22821)  
<223> megAI; SEQ ID NO: 13= translated amino acid sequence

<221> misc\_feature  
<222> (12505)...(13470)  
<223> megAI, AT-L

<221> misc\_feature  
<222> (13576)...(13791)  
<223> megAI, ACP-L

<221> misc\_feature  
<222> (13849)...(15126)  
<223> megAI, KS1

<221> misc\_feature  
<222> (15427)...(16476)  
<223> megAI, AT1

<221> misc\_feature  
<222> (17155)...(17694)  
<223> megAI, KR1

<221> misc\_feature  
<222> (17947)...(18207)  
<223> megAI, ACP1

<221> misc\_feature  
<222> (18268)...(19548)  
<223> megAI, KS2

<221> misc\_feature

<222> (19876)...(20910)  
<223> megAI, AT2

<221> misc\_feature  
<222> (21517)...(22053)  
<223> megAI, KR2

<221> misc\_feature  
<222> (22318)...(22575)  
<223> megAI, ACP2

<221> CDS  
<222> (22867)...(33555)  
<223> megAII; SEQ ID NO: 14= translated amino acid sequence

<221> misc\_feature  
<222> (22957)...(24237)  
<223> megAII, KS3

<221> misc\_feature  
<222> (24544)...(25581)  
<223> megAII, AT3

<221> misc\_feature  
<222> (26230)...(26733)  
<223> megAII, KR3 (inactive)

<221> misc\_feature  
<222> (26998)...(27258)  
<223> megAII, ACP3

<221> misc\_feature  
<222> (27393)...(28590)  
<223> megAII, KS4

<221> misc\_feature  
<222> (28897)...(29931)  
<223> megAII, AT4

<221> misc\_feature  
<222> (29953)...(30477)  
<223> megAII, DH4

<221> misc\_feature  
<222> (31396)...(32244)  
<223> megAII, ER4

<221> misc\_feature  
<222> (32257)...(32799)  
<223> megAII, KR4

<221> misc\_feature  
<222> (33052)...(33312)  
<223> megAII, ACP4

<221> CDS  
<222> (33666)...(43271)  
<223> megAIII; SEQ ID NO: 15= translated amino acid sequence

<221> misc\_feature  
<222> (33780)...(35027)

```

<223> megAIII, KS5

<221> misc_feature
<222> (35385)...(36419)
<223> megAIII, AT5

<221> misc_feature
<222> (37068)...(37604)
<223> megAIII, KR5

<221> misc_feature
<222> (37860)...(38120)
<223> megAIII, ACP5

<221> misc_feature
<222> (38187)...(39470)
<223> megAIII, KS6

<221> misc_feature
<222> (39795)...(40811)
<223> megAIII, AT6

<221> misc_feature
<222> (41406)...(41936)
<223> megAIII, KR6

<221> misc_feature
<222> (42168)...(42425)
<223> megAIII, ACP6

<221> misc_feature
<222> (42585)...(43271)
<223> megAIII, TE

<221> CDS
<222> (43268)...(44344)
<223> megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
      SEQ ID NO: 16= translated amino acid sequence

<221> CDS
<222> (44355)...(45623)
<223> megCIII, desosaminyl transferase, desosamine glycosyltransferase;
      SEQ ID NO: 17= translated amino acid sequence

<221> CDS
<222> (45620)...(46591)
<223> megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
      TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
      SEQ ID NO: 18= translated amino acid sequence

<221> CDS
<222> (46660)...(47403)
<223> megH, TEII; SEQ ID NO: 19= translated amino acid sequence

<221> CDS
<222> (47411)...(47980)
<223> megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence

<400> 1
ctcgagecga tgctcggcgg cgcggtgggc caaccagtcg tggacgtcgt cggtggcggt      60
gggaggtccg ccgtgccgag tcaggaaacg tattgccgat tgtgtggatt ccggagtcgc      120

```

atgaccgttg	acccgatccc	ccatacgcc	ctcccgtgat	gtcgtggg	gtccgtgcg	180
taccgccc	actgacatt	gtcgaatca	accccgccca	gtgtaggg	ccgcccgcga	240
cgggagaag	tccgtcgaac	aacttccgg	tgaccggtcg	ccggcgtcg	tgaaacgggc	300
gtcggagcac	ccgatcatt	ctgtcgggtg	acttcctaac	tgtcggcg	cacatctttc	360
tgaccgggtg	gttccgtgg	atgacgcgtt	cccggcccgt	ctggaactgt	gcgtgggact	420
gaccggttgc	ggcgtgtttt	cgcccgtttc	cgaactgcg	attcgtcgat	cgcgaggtg	480
ggagcgggtg	gctgaccggg	atgatctgca	atcatggcg	tcaatgacga	tctctttag	540
catgggtccg	gccgaggg	cgacaggccc	gaaacgccc	gcattccagc	tggttcgacga	600
cgtcgacatc	accgtgcaag	ccgcgatgac	accgacacca	cgccatgctg	gtgccgcact	660
ggaagggtg	cgcgatcagg	gaaatggccg	tgctactaga	cagacgcca	acagctgtcc	720
gggcctgcg	aaacagcatc	gatctgcgtc	agccgttcat	tgccccggcg	gcaccgcctt	780
ggaaatccgt	gccaccgg	gtccgcagtg	acgatcgcg	acccgggttt	cgagacagca	840
ggtagtaggc	gatgcaggcg	tttcgtctcg	cgccggacgc	gtcgactag	gtggaatccg	900
tcacagtctt	caatccggga	gcgttctatg	gcagttggcg	atcgaaggcg	gctgggccc	960
gagttgcaga	tggcccgggg	tctctactgg	gggttcgg	ccaacggcga	tctgtactcg	1020
atgctectgt	ccggacggga	cgacgacccc	tggacctggt	acgaacggtt	gcgggcccgc	1080
ggacggggac	cgtacgccag	tcgggcccga	acgtgggtg	tcggtgacca	ccggaccgcc	1140
gccgaggtgc	tcgccgatcc	gggcttcacc	cacggcccgc	ccgacgctgc	ccggtggatg	1200
caggtggccc	actgcccggc	ggcctcctgg	gccggcccct	tcggggagtt	ctacgcccgc	1260
accgaggacg	cggcgtcggt	gacagtggac	gccgactggc	tcagcagcg	gtgcgccagg	1320
ctggtgaccg	agctgggg	gcgcttcgat	ctcgtgaacg	acttcgccc	ggaggtccc	1380
gtgctggcgc	tcggtaaccg	gcccgcactc	aagggcgtg	accccgaccg	tctccgggtc	1440
tggacctcgg	cgacccgggt	atgcctggac	gcccaggtea	gcccgcaca	gctcgcggtg	1500
accgaacagg	cgtgaccgc	cctcgacgag	atcgacgcg	tcaccggcgg	tcgggacgcc	1560
gcggtgctgg	tgggggtgg	ggcggagctg	gcggccaaca	cggtgggcaa	cgccgtcctg	1620
gccgtcaccg	agcttcccga	actggcggca	cgacttgccg	acgaccggga	gaccgcgacc	1680
cgtgtggtga	cggaggtgtc	gcggacgagt	cccggcgctc	acctggaacg	ccgcaccgcc	1740
gcgtcggacc	gccgggtggg	cggggtcgac	gtcccgaacc	gtggcgaggt	gacagtggtc	1800
gtcgccgcg	cgaaccgtga	tcccgaagtc	ttcaccgatc	ccgaccggtt	cgacgtggac	1860
cgtggcggcg	acgccgagat	cctgtcgtcc	cgcccgggt	cgccccgcac	cgacctcgac	1920
gccctgggtg	ccacctggc	cacggcgggc	ctgcggggcg	ccgcgcgggt	gttgccccgg	1980
ctgtcccgtt	ccgggcccgt	gatcagacga	cgtcgggtcac	ccgtcgccc	tggtctcagc	2040
cgttgcccgg	tcgagctgta	gaggaagaac	gatgcgcgtc	gtgttttcat	cgatggctgt	2100
caacagccat	ctgttcgggc	tggtcccgt	cgcaagcgcc	ttccaggcgg	ccggacacga	2160
ggtacgggtc	gtcgccctcg	cgccctgac	cgacgacgtc	accggtgccc	gtctgaccgc	2220
cgtgcccgtc	ggtgacgacg	tggaaacttg	ggagtggcac	gcccacgcg	gccaggacat	2280
cgtcgagtac	atgcggaccc	tcgactgggt	cgaccagagc	cacaccacca	tgctctggga	2340
cgacctcctg	ggcatgcaga	ccaccttcac	cccgaacctt	ttcgccctga	tgagccccga	2400
ctcgctcatc	gacgggatgg	tcgagttctg	ccgctcctgg	cgtcccgaact	ggatcgctctg	2460
ggagccgctg	accttcgccc	ccccgatcgc	ggcccgggtc	accggaacct	cgcacgccc	2520
gatgctgtgg	ggtccggacg	tcgccaccgc	ggcccggcag	agcttccctg	gactgctggc	2580
ccaccaggag	gtggagcacc	gggaggatcc	gctggccgag	tggttcgact	ggacgctgcg	2640
gcgcttcggc	gacgaccgc	acctgagctt	cgacgaggaa	ctggtgctgg	ggcagtggac	2700
cgtggacccc	atccccgagc	cgtgcgggat	cgacaccggc	gtccggacgg	tgggcatgcg	2760
gtacgtcccc	tacaacggcc	cctcgggtgg	gcccgcctgg	ctgttgccgg	aaccggaacg	2820
tcggcgggtc	tgcctgaccc	tcggcggttc	cagccgggaa	cacggcatcg	ggcaggtctc	2880
catcggcgag	atgttgagcg	ccatcgccga	catcgacgcc	gagttcgtgg	ccaccttcga	2940
cgaccagcag	ttggtcggcg	tgggcagcgt	tcgggcaaac	gtccgtaccg	ccgggttcgt	3000
gccgatgaac	gtcctgctgc	ccacctgcgc	ggccaccgtg	caccacggcg	gcaccggcag	3060
ttggctgacc	gccgccatcc	acggcgatcc	gcagatcatc	ctctcggacg	ccgacaccga	3120
ggtgcacgcc	aagcagctcc	aggacctcgg	cgcggggctg	tcgctcccgg	tcgcggggat	3180
gaccgccgag	cacctgcgtg	gggcgatcga	gcgggttctc	gacgagccgg	cgtaccgcct	3240
cgggtgcggag	cggatgcggg	acgggatgcg	gaccgaccgc	tcgccggccc	aggtggctcg	3300
catctgtcag	gacctggccg	ccgaccgggc	ggcacgcggc	aggcagccgc	gtcgaaccgc	3360
cgagccgcac	ctgccgcgat	gacttccacc	accaccggga	ccggctgatg	ccggtcccgg	3420
aatccacacg	ccgactttcc	ttctgacacg	agggggcccc	ggtggttacc	tccaccaact	3480
tggacacgac	agcacggccg	gcactgaact	cgttgaccgg	gatgcgggtc	gtcgccgcct	3540
tcctggtctt	cttcacgcac	gtcctgtcga	ggctcatccc	gaacagctac	gtgtacgccg	3600
acggcctgga	cgcttcttgg	cagaccaccg	gacgggtggg	ggtgtcgttc	ttctttatct	3660
tcagcgggtt	cgtgctgacc	tggtcggcgc	gggccagcga	ctcgggtgtg	tcgttctggc	3720
gcagacgggt	ctgcaagctc	ttccccaacc	acctggtcac	cgcttcgcc	gccgtggtgt	3780



tgttcctggt	caccgggacg	gcggtgagcg	gtgaggcgct	gateccgaac	ctcctgctga	3840
tccacgcctg	gttcccggcc	ctggagatct	ccttcggcat	caacccggtg	agctggctcg	3900
tggcctgcga	ggcggtcttc	tacctgtgct	tcccgtctgt	cctgttctgg	atctccggtg	3960
tccgcccggg	gcggctgtgg	gcctgggccc	ccgtgggtgt	cgccgcgac	tgggcggtac	4020
cggtggctgc	cgacctcctg	ctgccgagtt	ccccgcgct	gateccgggg	cttgagtact	4080
ccgccatcca	ggactgggtc	ctctacacct	tccctgcgac	gcggagcctg	gagttcatcc	4140
tcgggatcat	cctggcccgc	atcctgatca	ccggtcgggt	gatcaacgtc	gggctgctcc	4200
ccgcgggtgt	gttggtcccg	gtcttcttcg	tcgcctcgct	cttccctgcc	gggtgtctacg	4260
ccatctcctc	gtcgatgatg	atccttcccc	tgggtctgat	catcgccagc	ggcgcgacgg	4320
ccgacctcca	gcagaagcgc	accttcatgc	gtaaccgggt	gatgggtgtg	ctcggcgacg	4380
tctccttcgc	gctctacatg	gtccacttcc	tgggtgatcg	ctacggggcg	gacctgctgg	4440
ggttcagcca	gaccgaggac	gccccgctgg	gtctcgcact	cttcatgac	attccgttcc	4500
tcgcgggtct	cctgggtgctg	tcgtggctgc	tgtacaggtt	cgtcgagcta	cccgtcatgc	4560
gtaactgggc	ccgcccggcc	tccgcccggc	gcaaaccgcg	cacggaaccc	gaacagaccc	4620
cttcccgcgc	gtaagaagga	cggtgcatcg	gtgaccacct	acgtctggtc	ctatctgttg	4680
gagtacgaga	gggaacgagc	cgacatcctc	gatgcgggtg	agaaggctct	cgccagtggc	4740
agcctgatcc	tcggtcagag	tgtggagaa	ttcgagaccg	agtacgcccg	ctaccacggg	4800
atcgcgcact	gcgtgggcgt	cgacaacggc	accaacgctg	tgaaactcgc	gctggagtcg	4860
gtaggtgtcg	gacgcgacga	cgaggctcgt	acgggtctcca	acaccgcccg	ccccacagtc	4920
ctggccatcg	acgagatcgg	cgcccggccg	gtcttcgtgg	acgtccgcga	cgaggactac	4980
ctcatggaca	ccgacctggt	ggaggcgggc	gtcaccgccg	gtaccaaggc	catcgtcccg	5040
gtgcacctgt	acgggcagtg	cgtggacatg	acagccctgc	gggaactggc	cgaccggcgg	5100
ggcctcaagc	tcgtggagga	ctgcgcccag	gcccacgggt	cccggcgggg	cggtcggctg	5160
gccgggacga	tgagcgacgc	ggcggccttc	tcgttctacc	cgacgaagg	cctcggcgcc	5220
tacggcgacg	gcggcgcggt	cgtcaccaac	gacgacgaga	cagcccgcgc	cctgcgacgg	5280
ctgcggtact	acgggatgga	ggagggtctac	tacgtcaccc	ggaccccggg	tcacaacagc	5340
cgcctcgacg	aggtgcaggc	cgagatcctg	cggcgcaaac	tgacccgggt	cgacgcgtac	5400
gtcgcggggt	ggcgggcggt	cgcccagcgg	tacgtcgacg	ggctcgccga	cctccaagac	5460
tcgcacggcc	tcgaactccc	agtggtcacc	gacggcaacg	aacacgtctt	ctacgtgtac	5520
gtcgtccgcc	acccgcgcgc	cgacgagatc	atcaagcgtc	tccgggacgg	gtacgacatc	5580
tccctgaaca	tcagctaccc	ctggccgggt	cacaccatga	ccggcttcgc	ccacctcggt	5640
gtcgcgtcgg	ggtcgctgcc	ggtcacccga	cggttgcccg	gcgagatctt	ctcccttccc	5700
atgtacccct	ccctccctca	cgacctgcag	gacaggggtg	tcgaggcggt	gcgggaggtc	5760
atcacccggc	tgtgacgagc	ccgcgtgtcg	tcagcgaaga	cccactctgg	aaggggccgg	5820
catgccgaac	agccactcga	ccacgtcgag	caccgacgtc	gccccgtacg	agcggggcga	5880
catctaccac	gacttctacc	acggccgtgg	caagggatac	cgtgccgaag	ccgacgcgct	5940
cgtggaggtc	gcccgcgaagc	acaccccaca	ggcggcgacc	ctgctggacg	tggcctgcgg	6000
gaccggatcc	cacctggtcg	agctggcgga	cagcttccgg	gaggtgggtg	gggtcgacct	6060
gtcggccgcc	atgctcgcca	ccgcgcgccg	caacgacccc	gggcgggaac	tgcaccaggg	6120
cgacatgcgc	gacttctccc	tcgaccgcag	gttcgacgtc	gtcacctgca	tgttcagctc	6180
caccggttac	ctcgtcgacg	aggccgaact	ggaccgtgcc	gtggcggaac	tggccgggtc	6240
cctcgcgcct	ggcggcaccc	tcgtcgtgga	gccttggtgg	ttcccggaga	cgttccggcc	6300
cggtcgggtc	ggggccgacc	tggtcaccag	cggtgaccgg	aggatctccc	ggatgtcgca	6360
caccgtcccc	gcgggtctgc	ccgaccgcac	cgcctcccgg	atgaccatcc	actacacggt	6420
ggggtcaccc	gaggccggga	tcgagcactt	caccgaggtg	cacgtgatga	ccctgttcgc	6480
ccgcgcgccg	tacgagcagg	ccttccagcg	ggcgggcctg	agctgctcgt	acgtcggcca	6540
cgacctgttc	tcgcccgggc	ttttcgtcgg	ggtcgcgcgc	gagccggggc	ggtaggggtc	6600
gaggagctgg	gcacgcaggg	ggtcttcacc	ttcaccgccg	agacgttcgc	cgacgagcgg	6660
ggggtgttcg	gcacggcgta	ccaggaggac	gtgttcgtgg	cggcgctcgg	ccgcccgcgt	6720
ttcccgggtg	cccaggtcag	caccacccgg	tcccggcggg	gtgtgggtccg	gggggtgcac	6780
ttcacgacga	tgcccggctc	catggcggaag	tacgtctact	gcgccagggg	tagggcgatg	6840
gacttcgccc	tcgacatccg	gcccgggttc	ccgaccttcg	gccggggccga	gccggtcgag	6900
ctctccgcgc	agtcgatggg	cgggctgtac	cttcccgtgg	gcacgggcca	cctgttcgtc	6960
tccctggagg	acgacaccac	cctcgtctac	ctgatgtccg	ccggttaact	ccccgacaag	7020
gaacggggcg	tgcaccccct	ggatccggag	ctggcggttc	cgatcccggc	cgacctcgac	7080
ctcgtcatgt	ccgagcgggg	ccgggtcgca	cccaccctcc	gggaggcccg	ggaccagggg	7140
atcctgcccg	actacgcgcg	ctgcccggcc	gccgcgcacc	gggtgggtgcg	gacgtgaccc	7200
cggccggggc	tgccggcccg	tgggtggtgt	cggcgcgctc	ggtttccctg	gttcggcggt	7260
cacccacgcc	ctggccgacc	tcccgggtgc	ggtgcggctc	gtcgcggggc	gggaggtcgt	7320
cgtgccctcc	ggtgccgctc	ccgactacga	gacgcaccgg	gtggacctca	ccgaacccgg	7380
agcgctcgcg	gaggtgggtc	cggacgcccg	ggcggtcttc	ccgttcgcgc	cccagatcag	7440

gggtacgtca	gggtggcgga	tcagcgagga	cgacgtggtc	gccgaacgga	cgaacgtcgg	7500
cctggtccgg	gacctgatcg	ccgtcctgtc	ccgtcgcgg	cacgccccgg	tggtggtctt	7560
cccgggcagc	aacacgcagg	tcggcagggg	caccgcccgg	cgggtcatcg	acggcagcga	7620
gcaggaccac	cccagggggc	tctacgacag	gcagaaacac	accggggaac	agctgctcaa	7680
ggaggccact	gcggccgggg	cgatccgggc	gaccagtctg	cggctgcccc	cgggtgttcgg	7740
ggtgcccggc	gccggcaccg	ccgacgaccg	gggggtggtc	tccaccatga	tccgtcgggc	7800
cctgaccggc	caaccgctga	cgatgtggca	cgacggcacc	gtccggcgtg	aactgctgta	7860
cgtgaccgac	gccgcccggg	ccttcgtcac	cgccctggac	cacgcccagc	cgctcgcggg	7920
acgccacttc	ctggtgggga	cggggcgttc	ctggccgctg	ggcgaggtct	tccaggcggg	7980
ctcgcgcagc	gtcgcggggc	acaccggcga	ggaccgggtg	ccgggtggtct	cgggtgccgc	8040
tccggcgcac	atggaccctg	cggacctgcg	cagcgtggag	gtcgaccccc	cccggttcac	8100
ggctgtcacc	gggtggcggg	ccacggtcac	gatggcggag	gcggtcgacc	ggacgggtggc	8160
ggcgttgggc	ccccgcccgg	ccgcccggcc	gtccgagccc	tcctgaccgg	ggtcaccggg	8220
gttcgtccta	cggcaccggc	ccgtcgcagc	ccgggtgccg	gaagatcgct	tcgagttccc	8280
ggagttcctc	ctcgcggcag	gtcagctcgg	cggcccgtaa	cgccgagtcg	agctgctcgg	8340
gtgtgcgggg	gccgatgaca	gcgcccagga	tcccggggcg	ggacaggacc	caggccagac	8400
cgacctcggc	cgggtccgcg	ccgaggcgtc	ggcagtagtc	ctcgtacgcc	tcgacgaggg	8460
ggcgtacggc	ggggaggagc	acctgggcgc	gtccctgcgc	cgacttgacg	gcggttccgg	8520
ctgccaaact	ctccagtacg	ccgtcgagca	gcccgcggcg	cagggggggac	caggcgaaca	8580
cgcccacccc	gtacgcctgg	gcggcggggc	ggacgtccag	ctcgggggtg	cggacggcca	8640
ggttggtacag	gcactggtgg	gagatcatgc	cgagcaggtt	gcggcgtgcc	gcgtctcctc	8700
gggcggcggc	gatgtgccag	cccgcacggg	tggaggagcc	gacgtaccgg	accttcccac	8760
tgccgaccag	atgttcggcg	gcctgccaca	cctcgtccca	cgggtgcggcg	cggtcgatgt	8820
ggtgcgtctg	gtagatgtcg	atgtggtcga	ccccgaggcg	gcggaggag	ttctcgcagg	8880
cggcgacgat	gtgtcggggc	gagagcccgc	cgctcgttag	ccgttcgctc	atctcgtctc	8940
ccaccttggt	cgccaggacg	gtctcctcgc	gtcgacctcc	gccctggggc	aaccaccgtc	9000
cgacgagttc	ctcgggtgtg	cccttgtaga	gccgccagcc	gtagatgtcg	gcgggtgtcg	9060
tgcagttgac	gccccgctcg	agggcgtggt	ccatcagccg	cagcgcgtcg	tcgtcgggtc	9120
cccgtccact	gaagttcacg	gtgccgagcc	agagtcggct	ggtgtgcaac	gccgatcgtc	9180
cgacgcgtac	ccgggcggac	ccggccccgg	tggttcccac	gtcgggtcacc	tgctcggcgcg	9240
gtgctggtgg	gcgagcgcct	ccagcacggg	tacgacctcg	gcggggggtcg	gcgcggccag	9300
cgctcctgc	cgagcttctc	cggcgttctc	ggcgtgggaa	cggtcctcga	ccactgtggc	9360
gagagcctgc	cagagggtgt	cggcgtcgac	ctcgtccgga	cggagggaaga	caccgcgtcc	9420
cagctcggcg	gtgcgctgac	cacgcaggac	acagtcaccac	tcgtgggcca	cggagatctg	9480
cggtagcggc	tggtgcagcg	cgggtggccca	gcttccggca	ccgcccgtgg	ggatgacggc	9540
ggcacagccc	ggcagcagga	tgttcatggg	aacgaagtcc	accaggcgga	cgttgtccgg	9600
caccgacgcc	ggatcgagcc	cggagcgggt	caccacgata	tcgcccgtcg	accgcgcgag	9660
ggtggccagt	gtccggagga	actcctgcgg	gttcgagggt	atgcccagcg	ccgagtatcc	9720
cccgggtgaag	cagacccggc	ggactccgtc	cgaggctcct	agccactgcg	gcacgacgga	9780
ggacccgttg	tagggcaaag	tccgggtgtg	caccgactcc	agtccggtct	ccaggcggaa	9840
gctctcgggg	agctgggtcg	cgctccactg	tccgacagcg	aggtcctcgc	tgtagtcgag	9900
gccgaaccgg	ccggcgacct	cggtagacca	gccgccgagc	gggtccggcc	ggtcgtcggc	9960
gggacgctgc	ccgcgcaggt	cctgggagcg	gctgcgggaag	tagccggtga	ggtcgtctgc	10020
ccacagcagc	cgggcgtggg	cggccccgca	ggccttgggc	gcgaccgccc	cggcgaagg	10080
gaagggtccc	cagagcacca	ggtcgggacg	ccagtccatg	gcgaactcga	cgagttcgtc	10140
gacgaaggag	tcgttggtga	ccaccgggaa	gacgaaccgg	gaggtggcct	cctcgatgcc	10200
gtgcaggaac	tcccacgagc	gcagttccgg	tccgcgtcgg	gcgaagtcca	ggtcgggtgg	10260
gtagcgggtg	acctgcgcgg	cggcctcagg	ggagatgtcg	aagagtcggg	ggtccgagcc	10320
gagtggcacc	gaggtcagtc	ccgcgcggac	gacgacgtcg	gtgagctcgg	gctgactggc	10380
caccgggacg	tcgtggccgg	cgggtgtgag	cgcccaggcc	agggggagca	ggccctggaa	10440
gtgggtacgg	tgcgcgaaac	aggtgagcag	gacccgcact	ggtcactcct	tggtcgagat	10500
gagggcgcca	acggtccggg	cgatgccctc	ggccagcggc	accggggggg	gccagccggg	10560
cagcgtccgg	aactcgggtg	agtcgaagtc	gtcgtcgcgg	aagtcgttgg	cctcggcggt	10620
ctccgggtga	gggacgctga	cgacggggc	cgcagggttg	ccgggtctgac	gtgccacgct	10680
ggcggcgacg	gtctcgaaga	tctcgcggag	gggtcggggc	tcgtccgcgc	tcggcggtcca	10740
gacgtcggcg	accagcgctc	cgtgggtgtg	cagtgcggcg	gtgaacgcgg	tggccacgtc	10800
ctcgacgtgc	aggaggttgc	ggcgcacgct	gccctcgtgc	cacatcgtga	tcggctcacc	10860
ggcgagggtc	cgccggatca	tggcgggtgac	gacaccccgg	ccgggtctgcc	ccgacggggc	10920
gctgtggccg	tagatcgcg	gcaggcgcag	gatcaccggg	tcgacgaccc	cgctcctcgg	10980
ggcctgacgc	aggatccgct	cggcctcgat	cttgtgctgg	gcgtaccggc	tggggggcgg	11040
gggggtcgcg	gcctgggtgg	tgctggcgaa	caggagcacc	ggcgcggggtc	cgggtcttgc	11100

ccgcagcgcg	gcgacgaggt	cgcgcatgat	gcccgcgttg	acgcgttcgg	cctcgggcac	11160
cgtggcgggc	ctgcgccagg	tcgacccgcc	ggcgggcgtag	gcgaccagat	gcacgacgac	11220
gtcgggtgctg	gcgacgacct	gcgcgacccg	gccgggttcg	agcaggtega	ctcgaagggtg	11280
ctcgatcccc	gcgctgcctg	gtggctgggtc	gcgagacccg	gtgcgcgcga	cggccccgag	11340
tcggagaggg	tgtgtggtaa	attcgcgaa	aagggcgctt	ccgacgaatc	cagaaacgcc	11400
gagaagtgtg	acatgtcttg	tcactacta	atgcattccg	atagccaccg	gcgcatggaa	11460
tccatttgtt	ccccccagg	tgggtgtcggg	tgacaaatcc	ggcctcaggt	cggcctcaag	11520
cctctttcga	gcgggtgctg	aggcttcccc	cgtaccctcg	gtggcctgcg	ttcggggcggg	11580
tgtcggggaa	agggcggtac	gaggagtctg	gtagggcgctc	gcggcgcgta	ctccgggact	11640
gatccgggtc	gacgccccga	cgcgtagacg	ggcgctcgatc	cgtgccgccc	gtaccgcccg	11700
ttttcggcga	tggtcgcaga	ttcctcccga	cgtgggtggac	tcattggttc	tcccgggtgt	11760
ggccgcaccg	tcgggtggcct	cgtcgggggt	gtcggagacc	gggtcgatcg	ccgtcccccg	11820
ccgtgccgac	cagggtcggt	ccgtcgccga	gggtgggtcac	cgtcgggtgg	acccgggtccg	11880
ccggcgccca	ccgcccgatc	gtgcccacct	tcgcctccgc	gggtaaatgc	ttcgtcgatc	11940
tgatcgacac	ttccggcgac	gctatcaccg	gagcattccc	cggcaccacc	ggtcgatgcc	12000
tcgcgctttc	caaacaggga	aaacagcagc	tcacagcggt	tccaggcgcc	gggcaatcct	12060
agcgaagagt	ctcgatgggg	tcaagggtgaa	ttctgtcaca	gatgtttttg	ttaaatgtac	12120
tttcttcagc	caccctcgac	gttcatacaa	ttggccggca	tctctaccaa	gggggagtga	12180
gtggttgacg	tgcccgatct	actcggcacc	cggactccgc	acccaggggc	gctcccattc	12240
ccgtggcccc	tgtgcggtca	caacgaaccg	gagctgcggg	cccgcgcccc	tcaattgcac	12300
gcatactctg	aaggcatttc	cgaggatgac	gtggtggccg	tcggcgccgc	cctcgcgcgc	12360
gagacacgcg	cgcaggacgg	gccgcaccgc	gccgtcgctg	tggcctcctc	ggtcaccgag	12420
ctgaccgccg	cgtcgcgcgc	cctcgcgccg	ggccgcccac	acccctcggt	ggtacgcggt	12480
gtcgcgccgac	ccacggcacc	ggtggtgttc	gtcctgcccg	gtcagggcgc	ccagtggccc	12540
ggcatggcga	cccgaactgt	cgccgagtcg	cccgtcttcg	ccgcggcgat	gcgggcctgc	12600
gagcgggctt	tcgacgaggt	caccgaactg	tcgttgaccg	aggtcctgga	ctcacccgag	12660
cacctgcgcc	gcgtcgaggt	ggtccagccc	gcgtctctcg	cgggtcgagac	ctcactggcc	12720
gccctgtggc	ggtcgttcgg	ggtgcgaccc	gacgcggtac	tcggacacag	catcggtgag	12780
ctggccgcgc	ccgaggtctg	cggcgccgctc	gacgtcgagg	ccgcgcgcgc	ggccgcgcgc	12840
ctgtggagcc	gcgagatggt	cccactggtg	ggccgggggtg	acatggcggc	ggtggcgctc	12900
tccccggccg	agctggcagc	ccgggtcgag	cgggtgggacg	acgacgtcgt	gccggccggg	12960
gtcaacggtc	cccggtcggt	gctgctcacc	ggcgtctccc	agcccatcgc	acggcggggtc	13020
gccgagctgg	cggcacaggg	cgtacgcgcc	caggtcgta	acgtgtcgat	ggcggcgcac	13080
tcggcgacag	tcgacgccgt	cgccgagggc	atgcgtctcg	cgtgacctg	gttcgcccc	13140
ggcgactccg	acgtgcctta	ctacgcgggc	ctcaccggcg	ggcggttgga	cacccgggaa	13200
ctcggcgccg	accactggcc	gcgcagtttc	cggctcccgg	tgcgtctcga	cgaggcgacc	13260
cgtgcgggtcc	tggaaactgca	gcccggcacg	ttcatcgagt	cgagcccgcga	cccgggtgctg	13320
gcggcctccc	tgcagcagac	cctcgacgag	gtcgggtccc	cggccgcgat	cgtgccgacc	13380
ctgcaacgcg	accagggcgg	tctgcggcgg	ttcctgctcg	ccgtggcgca	ggcgtaacc	13440
ggtggcggtga	cagtcgactg	gaccgcgcgc	taccccgggg	tgacccccgg	ccacctgccg	13500
tcggccgctg	ccgtcgagac	cgacgagggg	ccctcgacgg	agttcgactg	ggccgcgcgc	13560
gaccacgtac	tgcgcgcgcg	gctgctggag	atcgtcggcg	ccgagacggc	cgcgctcgcc	13620
gggcgggagg	tcgacgcccg	ggccaccttc	cgggaactgg	gcctcgactc	ggtcctcgcg	13680
gtgcagctgc	ggacccgcct	cgccacggcg	accgggcggg	atctgcacat	cgccatgctc	13740
tacgaccacc	cgaccccgcga	cgccctcacc	gagggcgctg	tgcgcggccc	gcaggaggag	13800
ccggggcggg	gtgaggagac	ggcacacccc	acggaggccg	aacccgacga	acccgtcgcc	13860
gtggtcgcca	tggcgtgccg	gctgcccggc	ggcgtcacct	caccggaggga	gttctggggg	13920
ctgctggccg	aggggcggga	cgcgctcggc	gggctgcccga	ccgaccgggg	atgggacctg	13980
gactcgctgt	tccaccccga	cccgaaccgg	tcgggcacgg	cgcaccagcg	cgtggtggc	14040
ttcctcaccg	gcgccacctc	cttcgacgct	gccttcttcg	ggctgtcgcc	acgggaggga	14100
ctggccgctg	agccgcagca	gcggatcacg	ttggagctgt	cgtgggaggt	gctggaacgc	14160
gccgggateg	ccccgacgtc	gttgccggacc	tcccggaccg	gggtgttcgt	cggctctgatc	14220
ccccaggagt	acggcccccg	gctggccgag	gggggtgagg	gcgtcgaggg	ctacctgatg	14280
accgggacca	ccaccagcgt	cgcctccggt	cgggtcgctt	acaccctcgg	cctggagggg	14340
ccggcgatca	gcgtcgacac	cgcctgctcg	tcgtcgctcg	tcgccgtgca	cctggcgctgc	14400
cagtcgctgc	ggcgcggcga	gtcgacgatg	gcgtcgcccg	gtggcggtgac	ggtgatgccg	14460
acaccgggca	tgctcggtga	cttcagtcgg	atgaactccc	tcgcccccca	cggacgggtcc	14520
aaggcgttct	cggccgcgcg	cgacgggttc	ggcatggccg	aaggcgaggg	gatgctcctg	14580
ctggaacggc	tctcggacgc	ccgccgccac	ggccaccccg	tgctcgccgt	gatcaggggc	14640
accgctgtca	actccgacgg	cgcgagcaac	ggactctccg	ccccgaacgg	ccgggcccag	14700
gtccgggtga	tccgacaggc	cctcgcggag	tccgggctga	cgccccacac	cgtcgacgtc	14760



gtggagaccc	acggcaccgg	cacccgcctc	ggtgatccga	tcgaggcacg	ggcgctctcc	14820
gacgcgtacg	gcggtgaccg	tgagcaccgg	ctgcggatcg	gctcgggtcaa	gtccaacatc	14880
gggcacaccc	aggccgccgc	cggtgtcgcc	ggtctgatca	aactgggtgtt	ggcgatgcag	14940
gccggtgtcc	tgccccgcac	cctgcacgcc	gacgagccgt	caccggagat	cgactgggtcc	15000
tcgggcgcga	tcagcctgct	ccaggagccc	gctgcctggc	ccgccggcga	gcggccccgc	15060
cgggccgggg	tgtcctcggt	cggcatacgc	ggcaccaacg	cacacgcgat	catcgaggag	15120
gcgcgcgcga	ccggtgacga	cacccgaccc	gaccggatgg	gcccgggtgg	gccctgggtg	15180
ctctcggcga	gcaccggcga	ggcgttgccg	gcccggggcg	cgcggtggc	cgggcaccta	15240
cgcgagcacc	ccgaccagga	cctggacgac	gtcgctact	cgctggccac	cggtcggggc	15300
gcgctggcgt	accgtagtgg	gttcgtgccc	gccgacgcgt	ccacggcgct	gcggatcctc	15360
gacgaactcg	ccgccgggtg	atccggggac	gcggtgaccg	gcaccgcccg	cgccccgcag	15420
cgcgctcgtc	tcgtcttccc	cggccaggga	tggcagtggg	cggggatggc	agtcgacctg	15480
ctcgacggcg	accgggtctt	cgccctcggtg	ctgcgggagt	gcgcgcgacg	ggttgaaccg	15540
tacctggact	tcgagatcgt	cccgttccctg	cgggccgagg	cgacgcgcgc	gacccccgac	15600
cacacgctct	ccaccgaccg	cgctcgacgtg	gtccagccgg	tgctgttcgc	ggtgatgggtg	15660
tccttggcgg	cccgggtggcg	ggcgtacggg	gtggaaccgg	cgcccgatcat	cggacactcc	15720
cagggggaga	ttgccgcggc	gtgtgtggcc	ggggcgctct	cgctggacga	cgcgggcccg	15780
gcggtggccc	tgcgagccg	ggtcatcgcc	accatgcccg	gcaacggcgc	gatggcctcg	15840
atcgccgcct	ccgtcgacga	ggtggcgggc	cggatcgacg	ggcggggtcga	gatcgccgcc	15900
gtcaacgggtc	cgcgcgcggt	ggtggtctcc	ggcgaccgtg	acgacctgga	ccgcctggtc	15960
gcctcctgca	ccgtcgaggg	ggtgcggggc	aagcggctgc	cggtggacta	cgcgctgcac	16020
tcctcgcacg	tcgaggccgt	ccgtgacgcg	ctccacgccg	aactcggcga	gttcggggcg	16080
ctgccgggct	tcgtgcggtt	ctactcgaca	gtcaccggcc	gctgggtcga	gcccggccga	16140
ctcgacgccg	ggtactgggt	tcgcaacctg	cgccacaggg	tcgggttcgc	cgacgcggtc	16200
cgctccctcg	ccgaccaggg	gtacacgacg	ttcctggagg	tcagcgccca	cccgggtgctc	16260
accacggcga	tcgaggagat	cggtgaggac	cgtggcggtg	acctcgctgc	tgtccactcg	16320
ctgcgacgtg	gggccggcgg	tcctcgctgac	ttcgggtccg	cgctggcccg	cgccctcggtg	16380
gccggcgctcg	cagtggactg	ggagtccggc	taccagggtg	ccggggcgcg	tcgggtgccc	16440
ctgcccacgt	accggttcca	gcgtgagcgc	ttctggttgg	aaccgaatcc	ggcccgcagg	16500
gtcgccgact	ccgacgacgt	ctcgctccctg	cggtaccgca	tcgaatggca	cccgaccgat	16560
ccgggtgagc	cgggacggct	cgacggcacc	tggctgctgg	cgacgtaccc	cggtcggggc	16620
gacgaccggg	tcgaggcggc	gcggcaggcg	ctggagtccg	ccggggcgcg	ggtcgaggac	16680
ctggtggtgg	agccccggac	gggccgggtc	gacctgggtg	ggcggtcga	cgccgtgggt	16740
ccggtggcgg	gcgtgctctg	cctgttcgct	gtcgcgagc	cggcggccga	acactccccg	16800
ctggcggtga	cgctggttgc	ggacacgctc	gacctgaccc	aggcggtggc	cggtcggggc	16860
cgggagtgtc	cgatctgggt	ggtcaccgag	aacgcctcg	ccgtcggggc	cttcgaacgg	16920
ctccgcgacc	cgccccacgg	cgcgctctgg	gccctcggtc	gggtcgctgc	cctggagAAC	16980
cccgcgctct	ggggcggcct	ggtcgacgtg	ccgtcgggtt	cggtcgccga	gctgtcgctg	17040
cacctcggga	cgacctgtc	cggcgcgggc	gaggaccagg	tcgcccctcg	acccgacggg	17100
acgtacgccc	gccggtgggtg	cagggcgggc	gcgggcggca	cgggccgggtg	gcagccccgg	17160
ggcacggtgc	tcgtcaccgg	cggcaccggc	ggggtcggtc	ggcacgtcgc	ccggtggctg	17220
gcccgcacgg	gcaccccggtg	cctggtgctg	gccagccgcc	ggggaccgga	cgccgacggg	17280
gtcgaggagc	tactcacgga	actcgccgac	ctgggcaccc	gggccaccgt	caccgcctgc	17340
gacgtcaccg	accgggagca	gctccgtgcc	ctcctcgcca	ccgtcgacga	cgagcaccgc	17400
ctgtcggcgg	tgttccacgt	cgccgcgacg	ctcgacgacg	gcaccgtcga	gacctcacc	17460
ggtgaccgca	tcgaacgggc	caaccggggc	aaggtgctcg	gtgcccgcaa	cctgcacgag	17520
ctgacccggg	acgccgacct	cgacgcgttc	gtgctcttct	cctcctccac	cgccgcgttc	17580
ggcgcgccgg	ggctcggcgg	ctacgtcccc	ggcaacgcct	acctcgacgg	tctcgcccag	17640
cagcgacgca	gcgagggact	cccggccacc	tcgggtggcg	gggttacctg	ggcgggcagc	17700
gggatggccg	agggtccggt	cgccgaccgg	ttccgcgggc	acggggatcat	ggagatgcac	17760
cccgaccagg	ccgtcgaggg	tctccgggtg	gcactgggtg	agggtgaggt	agccccgatc	17820
gtcgtcgaca	tcaggtggga	ccggttccct	ctcgcgtaaa	ccgcgcgacg	ccccacccgg	17880
ctcttcgaca	ccctcgacga	ggcccgtcgg	gccgcgcccg	gtcccgaacg	cgggccgggg	17940
gtggcgggcg	tggccgggct	gcccgtcggg	gaacgcgaga	aggcggtcct	cgacctggtg	18000
cggacgcacg	cggtgcccgt	cctcgccacc	gcctcgcccg	agcaggtgcc	cgtcgacagg	18060
gccttcgccc	aactcggcgt	cgactcgctg	tcggccctgg	aactgcgcaa	ccggctgacc	18120
actgcgaccg	gggtccggct	ggccacgacg	acggtcttcg	accacccgga	cgtacggacc	18180
ctggccggac	acctggccgc	cgaactgggc	ggcggtatcg	ggcgggagcg	gcccgggggc	18240
gaggccccga	cggtggcccc	gaccgacgag	ccgatcgcca	tcgtcgggat	ggcctgcccg	18300
ctgccggggg	gagtggactc	accggagcag	ctgtgggagt	tgatcgctct	cgggcgggac	18360
accgcctcgg	cggcacccgg	ggaccggagc	tgggatccgg	cggagttgat	ggtctccgac	18420

acgacgggca	cccgtaccgc	cttcggcaac	ttcatgcccg	gggcgggcga	gttcgacgcg	18480
gcgttcttcg	ggatctcgcc	gcgtgaggcg	ttggcgatgg	atccgcagca	gcggcacgcc	18540
ctggagacca	cctgggaggc	gctggagaac	gccggtatcc	ggcccagatc	gttgcgcggt	18600
acggacaccg	gtgtcttcgt	gggcatgtcc	catcaggggt	acgccaccgg	ccgcccgaag	18660
cccgaggacg	aggtcgacgg	ctacctgttg	acaggcaaca	ccgcgagcgt	cgccctccggt	18720
cggatcgcggt	acgtgttggg	gttggagggg	ccggcgatca	ctgtggacac	ggcgtgttcg	18780
tcgtcgcttg	tggcggtgca	cgtggcgggc	ggttcggttg	gttctgggga	ctgtggtctg	18840
gcggtggcgg	gtggggtgtc	ggtgatggcc	ggtccggagg	tgttcaggga	gttctcccgg	18900
cagggcgcgt	tggctccgga	cggcagggtg	aagcccttct	cggacgaggc	cgacggcttc	18960
ggtctggggg	aggggtcggc	cttcgctcgt	ttgcagcggg	tgtcgggtgg	ggtgcgggag	19020
gggctgcggg	tgttgggtgt	ggtggtgggt	tcggcggtga	atcaggatgg	ggcgagtaat	19080
gggttggcgg	cgccgtcggg	ggtggcgag	cagcgggtga	ttcggcgggc	gtggggtcgt	19140
gcgggtgtgt	cgggtgggga	tgtgggtgtg	gtggaggcgc	atgggacggg	gacgcggttg	19200
ggggatccgg	tggagtggg	ggcgttgttg	gggacgtatg	gggtgggtcg	gggtggggtg	19260
ggtccggtgg	tgggtgggtc	ggtgaaggcg	aatgtgggtc	atgtgcaggc	ggcggcgggg	19320
gtggtgggtg	tgatcaaggt	ggtgttgggg	ttgggtcggg	ggttgggtgg	tccgatggtg	19380
tgtcgggggt	ggttgcggg	gttgggtggat	tggctcgtcg	gtgggttggg	ggtggcggtt	19440
ggggtgcggg	ggtggccggt	gggtgtggat	ggggtgcgtc	ggggtggggg	gtcggcggtt	19500
ggggtgtcgg	ggacgaatgc	tcattgtggg	gtggcgagg	cgccgggggt	ggtggtgggg	19560
gcggaacggc	cgggtggagg	gtcgtcgcgg	gggttgggtg	gggtggttgg	tgggtgtggt	19620
ccggtggtgc	tgtcggcaaa	gaccgaaacc	gccctgcacg	cccaggcacg	tcgactcgcc	19680
gaccacctgg	agacgcaccc	cgacgtcccg	atgaccgacg	tgggtgtggg	gctgacgcag	19740
gcccgcaca	gcttcgacag	gcgcgcgggt	ctcctcgccg	ccgaccggac	ccaggccgtg	19800
gaacggctgc	gcggcctcgc	cggggggcgaa	ccggggaccg	gtgtggtgtc	gggggtggcg	19860
tcgggtggtg	gtgtggtgtt	tgtttttcct	ggtcaggggt	gtcagtgggt	ggggatggcg	19920
cgggggttgt	tgtcgggttc	ggtgtttgtg	gagtcgggtg	tggagtgtga	tgcgggtggt	19980
tcgtcgggtg	tggggttttc	ggtgttgggg	gtgttggagg	gtcggtcggg	tgcgccgtcg	20040
ttggatcggg	tggatgtggt	gcagccgggt	ttgttcgtgg	tgatggtgtc	gttggcgcg	20100
ttgtggcggt	ggtgtggggg	tgtgcctgcg	gcggtggtgg	gtcattcgca	gggggagatc	20160
gcggcgggcg	tgggtggcg	ggtgttgtcg	gtgggtgatg	gtgcgcgggt	ggtggcggtg	20220
cgggcgcggg	cgttgcgggc	gttggccggc	cacggcgcca	tggcctcggt	acgccgaggc	20280
cgcgacgacg	tacagaagct	cctcgacagc	ggccctgga	cggggaagct	ggagatcgcc	20340
gcggtcaacg	gccccgacgc	ggtggtggtc	tccggcgacc	cccagaccgt	gaccgagctg	20400
gtcgagcact	gtgacgggat	cggggtccgg	gcccggacga	tccccgtcga	ctacgcctcc	20460
cactccgcac	aggtcgagtc	gctccgggag	gagctgctct	ccgtcctggc	cgggatcgag	20520
ggccgccccg	cgacggtgcc	gttctactcc	accctcaccg	gtgggttcgt	cgacggcacc	20580
gaactggacg	ccgactactg	gtaccgcaac	ctgcgccacc	cgggtgcggt	ccacgccgcc	20640
gtcgaggcgc	tggcagcgcg	tgacctcacc	acgttcgtcg	aggtcagccc	gcaccccgtg	20700
ctgtcgatgg	cggtcggggg	gacgcttgcc	gacgtggagt	ccgccgtcac	tgtgggcacc	20760
ctggaacgcg	acaccgacga	cgtcgagcgc	ttcctcacct	ccctcgccga	ggcgcacgtc	20820
cacggcgtag	ccgtggactg	ggcggcggtc	ctcggctccg	gaaccctggt	cgacctgcc	20880
acctatccct	tccagggacg	gcggttctgg	ctgcaccccg	accgtgggtc	gcgtgacgat	20940
gtcgccgact	ggttccaccg	ggtcgactgg	acggcgacgg	ccaccgacgg	gtcggcccga	21000
ctcgacgggt	gctggtggt	ggtcgtaccc	gaggggtaca	cggacgacgg	ctgggtcgtg	21060
gaggtgcggg	ccgccctcgc	cgccgggtgg	gccgagccgg	tgggtgacgac	ggtcgaggag	21120
gtcaccgacc	gggtcggtga	cagcgacgcg	gtggtgtcga	tgtcggggt	ggccgacgac	21180
ggtgcggccg	agaccctggc	gctgctgcga	cgactcgacg	cacaggcgtc	caccacccca	21240
ctgtgggtgg	tcaccgtggg	ggcgtcgcc	cccgcgggtc	cgggtgcagcg	ccccgaacag	21300
gcgacgggtg	gggggttggc	ccttgtcgcc	tccctggaac	gcggacaccg	gtggaccggc	21360
ctgctggatc	tgccgcagac	accggacccg	cagctacgac	cccggctggg	cgaggcgctc	21420
gccggtgccg	aggaccaggt	agcggtcgcg	gccgacgccg	tacacgcccg	tcggatcgtc	21480
cccaccccg	tcaccggagc	cgggcccgtac	accgcccggg	gcgggacgat	cctcgtcacc	21540
gggggcaccg	ccggtctggg	tgccgtcacc	gcccgatggc	tcgccgagcg	cgggtgccga	21600
cacctcgccc	tggtcagccg	gcgcggggcg	ggcaccggcg	gcgtcgacga	ggtggtccgg	21660
gacctgaccg	ggctcggcgt	acgggtgtcg	gtgcactcct	gcgacgtcgg	cgaccgcgag	21720
tcggtcggcg	ccctggtgca	ggagtggaca	gcagccgggt	acgtgggtccg	gggggtggtc	21780
cacgctgccg	gtctgcccc	gcaggtgcca	ctgaccgaca	tggacccggc	cgacctcgcc	21840
gacgtggtgg	ccgtgaaggt	cgacggcgcg	gtgcacctgg	ccgacctgtg	cccggaggcc	21900
gaactgttcc	tgtgttctc	ctccggggcc	gggggtgtgg	gcagtgcccg	tcagggtgcg	21960
tacgcccgcc	gaaacgcctt	cctggacgcc	ttcgcccagc	accggcgggg	ccgggggtctg	22020
cccgccacct	cgggtggcgtg	ggggctctgg	gcggccgggg	ggatgacagg	ggaccaggag	22080

gcggtgtcgt	tcctgcgtga	gcggggcgta	cggccgatgt	cgggtgccgag	ggcactggaa	22140
gcgctggaac	gggtcctcac	cgccggggag	accgcggtgg	tcgtcgccga	cgtcgactgg	22200
gcggccttcg	ccgagtcgta	cacctccgcc	cggccccggc	cgctgctcca	ccggctcgtc	22260
acacctgcgg	cggcggtcgg	cgagcgcgac	gagccgcgtg	agcagacctt	ccgggaccgg	22320
ctggcgggccc	tgccccgggc	cgagcggtcg	gcggagctgg	tacgcctggt	ccggcgggac	22380
gccgcagccg	tgctcggcag	cgacgcgaag	gccgtaccgg	ccaccacgcc	gttcaaggac	22440
ctcgggttcg	actcgttggc	cgcggtccgg	ttccgtaacc	ggctggccgc	ccacaccggt	22500
ctgcgtctgc	cggccaccct	ggtcttcgag	caccggaacg	ccgcagccgt	cgccgacctc	22560
ctccacgacc	gactcggcga	ggccggcgag	ccgacccccg	tccggtcggt	gggcgcccga	22620
ctggccgcgc	tggagcaggc	cctgccccgac	gcctccgaca	cggagcgggt	cgagctggtc	22680
gagcgcctgg	aacggatgct	cgccggggctc	cgccccgagg	ccggagccgg	ggccgacgcc	22740
ccgaccgccg	gtgacgacct	gggggaggcc	ggcgtcgacg	aactcctcga	cgcgctcgaa	22800
cgggaactcg	acgccagggtg	aaccggaact	gaccgcagcc	gcagccgaag	cagagaccga	22860
ggacctgtga	ctgacaacga	caagggtggcg	gagtaacctc	gtcgtgcgac	gctcgacctg	22920
cgggcccggc	gcaagcgcct	gcgcgagctg	caatccgacc	cgatcgcggt	cgtcggcatg	22980
gcctgcgcgc	taccgggcgg	ggtgcacctc	ccgcagcacc	tgtgggacct	cctgcgccag	23040
gggcacgaga	cgggtgtccac	cttccccacc	gggcgcggct	gggacctggc	cgggctcttc	23100
caccgggacc	ccgaccaccc	cggcaccagc	tacgtcgacc	ggggtgggtt	cctcgacgac	23160
gtggcgggct	tcgacgccga	gttcttcggg	atctccccgc	gcgaggccac	ggccatggac	23220
ccgcaacagc	ggctgctggt	ggagaccagt	tgggagctgg	tggagagcgc	cggcatcgat	23280
ccgcactccc	tgctgtggac	cccgaccggc	gtcttcctcg	gcgtggcgcg	gctcggtac	23340
ggcgagaacg	gcaccgaagc	cggtgacgcc	gagggtattt	cggtgaccgg	ggtggcacc	23400
gctgtgcctt	ccgggcggat	ctcctacgcc	ctcgggctgg	agggctccgt	gatcagcggt	23460
gacaccgcgt	gctcgtcgtc	gttgggtggcg	ctgcacctgg	cggtcgagtc	gctgcggctg	23520
ggcgagtcga	gtctcgtgtt	cgtcggcggg	gcggcggtca	tggcgacacc	aggggtgttc	23580
gtcgacttca	gccgccagcg	ggcgttggcc	gctgacggca	ggtcgaaggc	cttcggggcc	23640
gccgccgacg	ggttcggctt	ctccgagggg	gtctccctcg	tcctgctcga	acggctctcc	23700
gaggccgaaa	gcaacggcca	cgagggtgtt	gctgtcatcc	gtggctccgc	cctcaaccag	23760
gacggggcca	gcaacggtct	cgccgcgcgc	aacgggaccg	cccagcgcaa	ggtgatccgg	23820
caggcgctac	gaaactgcgg	cctgaccccc	gccgacgtgg	acgccgtgga	ggcgcacggc	23880
accggcacca	cgtcggcgga	cccgatcgag	gccaacgccc	tgctggacac	ctacggccgt	23940
gaccgggata	cggaccaccc	gctgtggctg	gggtcgggtg	agtccaacat	cggccacacg	24000
caggcgggcg	cgggcgtcac	cgggctgctc	aagatgggtg	tggcactgcg	ccacgaggaa	24060
ctgcccggca	ccttgacgtt	cgacgagccc	accccgacag	tggactgggt	ctcgggagcg	24120
gtacgcctgg	cgacccgggg	ccggccgtgg	cggcggggtg	accggccgag	gcgggcccgg	24180
gtgtcggcgt	tcggcatcag	cgggaccaac	gcccacgtga	tcgtcgagga	ggcacccgag	24240
cggaccaccg	agcgcaccgt	cggcggcgac	gtcggcccgg	tcccgtcgtt	ggtgtccgcc	24300
cggtcggcgg	cggcgctacg	ggcccaggcg	gcccaggctg	ccgagctggt	ggagggctcc	24360
gacgtcgggc	tggcggaggt	cgggcgggagc	ctggccgtga	cccgggcgcg	acacgagcac	24420
cgggcggcgg	tgggtggcgt	gacccggggc	gaggcggtgc	gggggctgcg	cgaggctcgc	24480
gcggtcgaac	cgcgcggcga	ggacaccgtc	accggggctg	ccgagacgtc	cgggcgcacc	24540
gtcgtcttcc	tcttcccggg	acaggggtcc	cagtgggtcg	ggatgggcgc	ggagctgctg	24600
gactcggcac	cggcggttcgc	cgacacgatc	cgcgcctgcg	acgaggcgat	ggcaccgttg	24660
caggactggt	cggctctccga	cgtgctccgg	caggagccgg	gggcaccggg	actggaccgg	24720
gtcgacgtgg	tgacgccggt	gctgttcgcg	gtgatgggtg	cgttggcgcg	gttgtggcag	24780
tcgtacgggg	tcacccccgc	tgccgtgggt	gggcactcgc	agggggagat	cgccgccggc	24840
cacgtggcgg	gtgcgtcttc	cctcgccgac	gcggcgaggc	tgggtgggtg	ccgcagccgg	24900
ttgctgcggt	cgtgttcggg	gggcggcggc	atgagcgccg	tcgcgctcgg	tgaggccgag	24960
gtacgccgcc	gactgcgggt	gtgggaggac	cggatctccg	tggccgccgt	caacggaccc	25020
cggtcgggtg	tgggtggccg	ggaaccggag	gcgctgcggg	agtggggacg	ggagcgggag	25080
gccgagggcg	tacgggtccg	cgagatcgac	gtcgactacg	cctcgacttc	gccgcagatc	25140
gacaggggtcc	gtgacgaact	cctgacgggt	acggggggaga	tcgagccccg	gtcggcgggag	25200
atcaccttct	actcgacgggt	cgacgtccgt	gctgtcgacg	gcaccgacct	ggacgcgggg	25260
tactggtacc	gcaacctgcg	ggagacgggt	cggttcggcg	acgcgatgac	ccggttggcc	25320
gactcgggat	acgacgcggt	cgtcgagggt	agcccgcatc	cgggtgggtg	gtcggcggtc	25380
gccgagggcg	tcgaggaggc	aggtgtcgag	gacgccgtcg	tcgtcggcac	cctgtcccgg	25440
ggcgacggcg	gaccgggggc	gttcctgcgg	tcggcgggca	ccgccacttg	cggcggtgtg	25500
gacgtcgact	ggacgcccgc	cctcccggga	gctgcgacga	tcccgttgcc	gacgtacccg	25560
ttccaacgga	agccgtactg	gctgcgggtc	tctgtctccg	ccccgccttc	ccacgatctc	25620
gcctaccggg	tgtcctggac	gccgatcacc	ccgcccgggg	acggcggtact	cgacggcgac	25680
tggctgggtg	tgacccccgg	gggcagcacc	ggatgggtcg	acgggttggc	ggcggcgac	25740



accgcccggcg	gtggccgggt	cgtegccac	ccggtggact	ccgtgaccte	ccggaccggc	25800
ctggccgagg	cgctcgcccg	gcgggacggc	acgttcggg	gggtgctgtc	gtgggtggcg	25860
accgacgaac	ggcacgtcga	ggccgggtgcg	gtcgccctgc	tgaccctggc	gcaggcggtg	25920
ggtagacggc	gaatcgacgc	accactgtgg	tgcctgaccc	aggaggcggt	ccgtaccccc	25980
gtcgacgggt	acctggccc	accggcgag	gccgccctgc	acggtttcgc	ccaggtcgcc	26040
cggtggagc	tggcccggcg	cttcgggtggg	gtgctcgacc	tgcccggcac	cgtecgacgc	26100
gccgggacgc	gtctggtcgc	ggcggtcctc	gccggcgggc	gcgaggacgt	cgtecgccgtc	26160
cgtggcgacc	gtctctacgg	ccgtcgccctg	gtcaggggcga	ccctggccgc	gcccggcggg	26220
gggttcaccc	cgcacggcac	cgteccgtgc	accggcgcg	ccggtcgggt	ggcggtcg	26280
ctggcccgggt	ggctcgccga	acgggggtgcc	acccgactcg	tcctggccc	cgcacacccg	26340
ggcgaggagt	tgtcgaccgc	gatccggg	gccgggtgcca	ccgccgtggt	gtgcgaaccg	26400
gaggcgagg	cactgcgtac	ggcgatcggc	ggggagtgc	cgaccgcgt	cgtacacgc	26460
gagacgttga	cgaacttcgc	cggcgtcgcc	gacgccgacc	ccgaggactt	cgccgccacc	26520
gtcgcgcgga	agaccgcgt	gccgacggtc	ctggcgagg	tgtcgggcga	ccaccgcctc	26580
gaacgggagg	tctactgtc	gtcggtggcc	ggggtctggg	gtggggtcgg	catggccgcg	26640
tacggccgcg	gcagcgcccta	cctcgacgcc	ctggtcgagc	accgtcgcg	ccgggggcac	26700
gccagcgcc	cggtggcctg	gaccccggtg	gccctgccc	gcgcggtcga	cgacggtcgg	26760
ctgcgcgagc	gcggcctgcg	cagcctcgac	gtggccgacg	ccctcgggac	gtgggaacgt	26820
ctgctccgcg	ccggtgcgg	gtcggtggcc	gtcgccgacg	tcgactggtc	ggtcttcaca	26880
gagggtttcg	cgcccatccg	gccgacccc	ctcttcgacg	aactcctcga	ccggcgcggg	26940
gaccccgacg	gcgcgcccgt	cgaccggccg	ggggagccgg	cgggcgagtg	gggtcgacga	27000
atcgcgcg	tgtcccgcga	ggaacagcg	gagacgttgc	tgaccctcgt	cggcgagacg	27060
gtcgcgagg	tgtgggaca	cgagaccggc	accgagatca	acacccgctc	ggccttcagc	27120
gaactcgcc	tcgactcgct	gggtcctgat	gccctgcgtc	agcgccctgg	ggcccgtacc	27180
ggcctgcgga	tgcggccctc	gctgggtctc	gaccaccgga	cggtcaccgc	gctcgcgcg	27240
tacctgcgtc	gactggtcgt	cggggactcc	gacccgaccc	cggtacgggt	gttcggcccc	27300
accgacgagg	ccgaacccgt	cgccgtggtc	ggcatcggt	gccggttccc	cggcggcatc	27360
gccaccccc	aggacctctg	gcgggtgggt	tcggagggca	cctccatcac	caccggattc	27420
cccaccgacc	ggggtggga	cctccggcg	ctctaccacc	ccgacccgga	ccaccccggc	27480
accagctacg	tcgacagggg	gggattcctc	gacggggccc	cgacttcga	ccccgggttc	27540
ttcgggatca	ccccccgcga	ggcgctggcg	atggaccgc	agcagcggt	caccctggag	27600
atcgcggtgg	aggcggtgga	acgggcgggc	atcgaccgg	agaccctcct	cggcagcgac	27660
accggcgctc	tcgtcgccat	gaacggccag	tcctacctgc	aactgctgac	cggggagggt	27720
gaccggctca	acggctacca	ggggttgggc	aactcggcga	gcgtgctctc	cgcccggtgc	27780
gcctacacct	tcgggtggga	ggggccggcg	ctgacgggtg	acaccgcctg	ctcgctcctc	27840
ctggtcgcca	tcacacctgc	catgcagtcg	ctgcgtcggg	gtgagtgtc	gctggcggtg	27900
gccggcgggg	tgacggtcac	ggccgacccg	tacaccttcg	tggacttcag	cgcacagcg	27960
gggctcgccg	ccgacggg	gtgcaaggcg	ttctccgcgc	aggccgacgg	gttcgccctc	28020
gccgaggggc	tcgcggcgt	cgtecccgaa	ccgttgcca	aggcgcgggc	aaacggccac	28080
cagggtgctg	cggtgctgcg	cggcagcgcc	gtcaaccagg	acggggccag	caacggcctc	28140
gccgccccga	acgggcccgtc	gcaggaacgg	gtgatcaggc	aggccctgac	cgcctccggg	28200
ctgcgtccc	ccgacgtcga	catggtggag	gcgcacggga	cgggcaccga	actcggcgac	28260
ccgatcgagg	ccggggcgct	catcgcgcg	tacggccggg	accgggaccg	gccgctctgg	28320
ctgggctcgg	tgaagacgaa	catcggccac	acccaggccg	ccgccgggtg	cgccgggggtg	28380
atcaaggcg	tcctggcgat	gcggcacggc	gtactcccga	ggtcgctgca	cgccgacgag	28440
ttgtccccgc	acatcgactg	ggcgacggg	aaggctcgagg	tgtctccgcga	ggcacgacag	28500
tggccccccg	gtgagcgccc	ccgcgcgc	ggggtgtcct	ccttcggcgt	cagcgggacc	28560
aacgcccacg	tcatcgtcga	ggaggcaccc	gccgaaccgg	accccgaaac	ggttcccgc	28620
gccccggg	ggccctgcc	cttcgtcctg	cacggacgca	gcgtccagac	ggtccggtec	28680
caggcgcgga	ccctcgccga	acacctgcgc	accaccggcc	accgggacct	cgccgacacc	28740
gcccgtaccc	tggccaccgg	tcgcgcgccgt	ttcgacgtcc	gggcgcag	gctcggcacc	28800
gaccgggagg	gtgtctgcgc	cgccctcgac	gcgctggcgc	aggatcgccc	ctcgcccgac	28860
gtcgtcgccc	cggcggtctt	cgccgcgccgt	accccgctcc	tggctctccc	cgggcagggg	28920
tcgcagtggg	tcggcatggc	ccgtgacctg	ctcgactcct	ccgagggtgt	cgccgagtcg	28980
atgggcccgt	gcgcgcaggc	gctgtcgccg	tacaccgact	gggacctgct	cgacgtggtc	29040
cgtggggctg	gcgaccccga	cccgtaacgac	cggtggacg	tgtccagcc	ggtgctgttc	29100
gcgggtgatg	tgtcgctggc	gcggttgtgg	cagtcgtacg	gggtgactcc	gggtgcgggtg	29160
gtgggtcact	cgcaggggga	gatcgccgcc	gcgcacgtgg	ctggtgcgtt	gtcgttggcc	29220
gacgccgcca	gggtggtggc	gttgccgcagc	cggtgctgc	gggagctcga	cgaccagggc	29280
ggcatggtgt	cggtcggcac	ctcccgcgc	gagttggact	cggtcctgcg	ccggtgggac	29340
gggcgggtcg	cggtggcggc	ggtgaacgga	cccggcacgc	tcgtgggtggc	cggaccacc	29400

gccgaactgg	acgagttcct	cgcggtggcc	gaggcccgcg	agatgaggcc	gcgtcggatc	29460
gcggtgcgct	acgcgtcgca	ctccccggag	gtggcccggg	tcgaacagcg	gctcgccgcc	29520
gaactcggca	ccgtcaccgc	cgtcggcggc	acggtcccgc	tctactccac	cgccaccggg	29580
gacctcctcg	acaccacagc	catggacgcc	gggtactggt	accgcaacct	gcgccaaccg	29640
gtgctgttcg	agcacgccgt	ccgcagcctc	ctggagcggg	gattcgagac	gttcattcgag	29700
gtcagcccgc	accctgtgct	gctgatggcg	gtcgaggaga	ccgccgagga	cgccgagcgc	29760
ccggtcaccg	gcgtgccgac	gctgcgccgc	gaccacgaag	ggccgtcgga	gttcctccgc	29820
aacctcctgg	gggcgcacgt	gcacggggtc	gacgtcgacc	tgcgtccggc	ggtcgcccac	29880
ggccgcctgg	tcgacctgcc	cacctacccc	ttcgacaggc	agcggctctg	gcccgaagccg	29940
caccgcaggg	ccgacacctc	gtcgctgggg	gtccgtgact	cgacccaccc	gctgctgcac	30000
gccgcagtcg	acgtaccccg	tcacggcgga	gcggtgttca	ccgggcggct	ctcccccgac	30060
gagcagcagt	ggctgaccca	gcacgtggtg	ggtgggcgga	acctggtgcc	cggcagtgtc	30120
ctggtcgacc	tcgcgctcac	cgccggggcc	gacgtcggcg	tgccggtgct	ggaggaactc	30180
gtcctgcagc	agccgctggt	gttgaccgcc	gccggtgcgt	tgtgcgcctc	gtcggtcggc	30240
gccgccgacg	aggacggggc	gcggccggtc	gagatccacg	ccgccgagga	cgtctccgac	30300
ccggccgagg	cccggtggtc	ggcgtagcgc	accgggaccc	tcgccgtcgg	cgtggccggc	30360
ggcgcccggg	acggcacaca	gtggcccccg	cccggcgcca	ccgccctgac	gttgaccgac	30420
cactacgaca	ccctcgccga	actgggctac	gagtaggggc	cggcgttcca	ggcgctgcgc	30480
gccgcgtggc	agcacggcga	cgtggtctac	gcggaggtgt	ccctcgacgc	cgtcgaggag	30540
gggtacgcgt	tcgaaccggg	gctgctcgac	gccgtcgccc	agaccttcgg	cctgaccagt	30600
cgcgcccccg	ggaagctccc	cttcgcctgg	cggggcgtca	ccctgcacgc	caccggggcc	30660
actgcggtac	gggtggtggc	gacccccgcc	ggaccggacg	cggtagccct	gcgggtcacc	30720
gacccgaccg	gtcagctcgt	cgccacgggt	gacgccctgg	tcgtcaggga	cgccggggcg	30780
gatcgggacc	agccgcgcgg	ccgcgacggc	gacctgcacc	gcctggagtg	ggtacggctg	30840
gccacccccg	acccgacccc	ggcgggcggt	gtgcacgtgg	cggccgacgg	gctcgacgac	30900
ctgctgcgcg	ccggtggtcc	ggcaccacag	gccgtcgctc	tcgcctaccg	tcccgaacggc	30960
gacgacccga	cggccgaggg	ccgtcacggg	gtgctctggg	cggccacgct	cgtgcgccgt	31020
tggctcgacg	acgaccggtg	gcccgcaccc	accctggtgg	tggccacgtc	cgagggggtc	31080
gaggtctccc	ccggggacga	cgtgccgcgc	cccgggggcg	ccgccgtgtg	gggggtgctg	31140
cgctgcgccc	aggcggagtc	cccgacccgc	ttcgtgctcg	tcgacggcga	cccggagacg	31200
cccccgggcg	tgccggacaa	tccgcagctc	gcggtccgtg	acggtgcggt	gttcgtgcca	31260
cggctgacgc	cgtcgcgcgg	tcccgtgccg	gccgtcgccg	accggggcgta	ccggctggtg	31320
cccggcaacg	gcggctccat	cgaggcagtg	gccttcgccc	ccgtcccccga	cgccgaccgg	31380
cccctggcgc	cggaggaggt	acgcgtcgcc	gtccgcgcca	ccggcgtgaa	cttcctgac	31440
gtcctgctcg	cgtcggcat	gtacccggaa	ccggccgaga	tgggcaccga	ggcgctccgt	31500
gtggtcaccg	aggtcgggtc	gggtgtccgg	cggttcaccc	ccggccaggc	ggtgacgggc	31560
ctgttccagg	gggccttcgg	gccggtggcg	gtcgccgacc	accggctcct	caccccggtc	31620
cccgcggggt	ggcggggcgg	ggacgccgca	gccgtaccca	tcgcgttcac	caccgcccac	31680
tacgcgtcgc	acgacctggc	cgggttgacg	gccgggcagt	ccgtgctggt	ccacgcccgc	31740
gccggcgggg	tggggatggc	tgccgtcgcg	ttggcccgtc	gggccggggc	ggaggtgttc	31800
gccacggcca	gcccggccaa	acacccgacg	ctgcggggcg	tcggcctcga	cgacgaccac	31860
atcgccctcg	cccgggagag	cgggttcggt	gagcgggttcg	ccgcgcgtac	cgggggggcg	31920
ggcgctcgacg	tggctcctgaa	ctcgctcacc	ggcgacctgc	tcgacgagtc	cgcgcggtcg	31980
ctcgccgacg	gcgggggtctt	cgtcgagatg	ggcaagaccg	acctgcggcc	ggcggagcag	32040
ttccggggcc	ggtacgtccc	gttcgacctg	gccgaggccg	gtcccgatcg	gctcggcgag	32100
atcctggagg	aggtcgtcgg	tctgctggcc	gccggtgccc	tcgaccggtt	gccggtgtcg	32160
gtgtgggagt	tgtcggcggc	cccggcccgcg	ctcacccaca	tgagccgggg	ccgacacgtg	32220
ggcaagctcg	tcctcaccca	gcccgcctcc	gtgcaccccg	acggaacggt	gctggtcacc	32280
ggcgggaccg	gcacctggg	gcggctggtc	gcccgcaccc	tggtagaccg	gcacggcgta	32340
ccccacctcc	tggtagccag	ccggcgcggt	ccggcgcccc	cgggcgcggc	cgagctgcgc	32400
gccgacgtcg	aaggcctcgg	cgcgaccatc	gagatcgctc	cctgcgacac	cgccgaccgg	32460
gaggcgctcg	cggcgctgct	cgactcgatc	cccgcggacc	gtccgctgac	cggggtggtg	32520
cacaccgccg	gggtcctggc	cgacgggctg	gtcacctcca	tcgacgggac	cgccaccgat	32580
caggctcctgc	gggccaagggt	cgacgcggcg	tggcacctgc	acgacctgac	ccgggacgcg	32640
gacctgagct	tcttcgtgct	gttctcgtcg	gcggcgctcg	tgtggccggg	tcccgggcag	32700
ggcggtgtacg	cggcggccaa	cggggtcctc	aacgccctgg	ccgggcaacg	gcggggccctc	32760
ggactgcccc	cgaaggcgct	cgggtggggc	ctgtggggcg	aggccagcga	gatgaccagc	32820
ggcctcggtg	accggatcgc	ccgtaccggg	gtcgccgcgc	tgcgaccga	gcgggcgctg	32880
gccctgttcg	acgcggctct	gcgcagcggc	ggggaggtgc	tgttcccgtc	gtctgtcgac	32940
aggtcggcg	tgcgcccggc	cgagtagctc	cccagaggtg	tgcgcggcgc	ggtccggctc	33000
acgccacggg	ccgccaacag	ggccgagacc	ccggggccggg	gcctgctcga	ccgtctcgtc	33060

ggtgcacccg	agaccgatca	ggtggccgcg	ctggccgagc	tggtecgctc	gcacgcggcg	33120
gcggtcgccg	gctacgactc	ggccgaccag	ctgcccgaac	gcaaggcggt	caaggacctc	33180
gggttcgact	cgctggcggc	ggtggagctg	cgcaaccggc	tcggcggtcac	caccggcgta	33240
cggtgcccc	gcacgctggt	gttcgaccac	ccgacaccgc	tggcggtggc	cgaacacctg	33300
cggtcgaggt	tgttcgcccga	ctccgcgcgc	gacgtcgggg	tcgggtgcgcg	cctcgacgac	33360
ctggaacggg	cgctcgacgc	cctgcccgcg	gcgcaggagc	acgccgacgt	cggggcccgc	33420
ctggaggcgc	tgctgcgcgc	gtggcagagc	cgacgacccc	cggagaccga	gccagtgcg	33480
atcagtgcg	acgccagtga	cgacgagctg	ttctcgatgc	tcgacaggcg	tctcggcggg	33540
ggaggggacg	tctaggtgac	aggtcgattc	cgccccgcgg	cagtggaccg	taccgccctg	33600
acagggtccac	cggttcgcgc	tcgcctccca	caccgcagcg	ccgggggtatc	cacggaaggg	33660
atccgatgag	cgagagcagc	ggcatgaccg	aggaccgcct	ccggcgctat	ctcaagcgca	33720
ccgtcgccga	actcgactcg	gtgacaggtc	ggctcgacga	ggtecgagta	cgggcccgcg	33780
aaccgatcgc	cgctcgtcggc	atggcctgcc	ggttcccccg	gggtgtggac	tcgccggagg	33840
cgttctggga	gttcateccg	gacggtggtg	acgcgatcgc	cgaggcgccc	acggaccgtg	33900
gctggccgcg	ggcaccgcga	ccccgcctcg	gtggtctcct	cgcgagaccg	ggcgcgttcg	33960
acgccgcctt	cttcggcatc	tcaccccgcg	aggcgctcgc	gacggacccc	cagcagcgcc	34020
tgatgctgga	gatctcctgg	gaggcggttg	agcgtgcggg	tttcgacccg	tcgagcctgc	34080
gcggcagcgc	cggtggcgct	ttcaccggtg	tcgggtgcgg	ggactacgga	cccaggccgg	34140
acgaggcacc	cgaggagggtg	ctcggttacg	tcggcatcgg	caccgcctcc	agcgtcgctt	34200
ccggacgggt	ggcgtaacac	ctgggggttg	agggtccagc	cgtaaccgtc	gacaccgcct	34260
gctcctccgg	gctcaccgcg	gtgcacctgg	cgatggagtc	gctgcgccgc	gacgagtgc	34320
ccctggctct	cgccggtggg	gtcaccgtga	tgagcagccc	gggtgcgttc	accgagttcc	34380
gcagccaggg	cggtgtggcc	gaggacggcc	gctgcaaacc	gttctcccgc	gccgccgacg	34440
gcttcgggct	cgccgagggg	gccgggggtcc	tgggtgctcca	acggctgtcc	gtcgcccggg	34500
ccgagggccg	gccggtgctg	gccgtactgc	gtggctcggc	gatcaaccag	gacggtgcca	34560
gcaacgggct	caccgcgcgc	agcggccccc	cccagcgggc	ggtgatcagg	caggcggttg	34620
agcgggcgcg	gctgcgtccc	gtcgacgtgg	actacgtgga	ggcccacggc	accggcaccc	34680
ggctgggcga	tcgatcgag	gcgcacgccc	tgctcgacac	gtacgggtgc	gaccgggaac	34740
ccggccgccc	gctctgggtc	ggatcggtga	agtcacaacat	cggtcacacc	caggcggcgg	34800
cggggggtgg	cggggtgatg	aagaccgtgc	tggcgctgcg	gcacgcggag	atcccggcga	34860
cgttgcactt	cgacgagccc	tcgccgcacg	tcgactggga	ccgggggtgcg	gtgtcggttg	34920
tgtccgagac	ccggccctgg	ccggtggggg	agcgcgcgcg	ccgggcgggg	gtgtcctcgt	34980
tcggcatcag	cggcaccaac	gcgcacgtca	tcgtcgagga	ggcgccgagc	ccgcaggcgg	35040
ccgacctcga	cccgaccccc	ggcccgcgaa	ccggagcgac	ccccggaacg	gatgccgccc	35100
ccaccgccga	gccgggtgcg	gaggcggtcg	cactgggtgtt	ctccgcgcgc	gacgagcggg	35160
ccctgcgcgc	ccaggcggcc	cggtcgcgcg	accgtctcac	cgacgacccc	gccccctcgt	35220
tgcgcgacac	cgccttcacc	ctggtcacc	gccgtgccac	ctgggagcat	cgggcggctc	35280
tcgtcgccgg	gggcgaggag	gtcctcgccg	gcctccgggc	cgtecgccgg	ggacgtcccc	35340
tcgacggagc	cgtcagcggg	cgggcgcgcg	ccggccgcgc	ggtggtgctg	gtcttccccg	35400
ggcaggggcg	acagtggcag	ggcatggccc	gggacctgct	gcggcagtcg	ccgaccttcg	35460
cgaggtccat	cgacgcctgc	gagcggggcg	tcgccccgca	cgtggactgg	tcgctgcgcg	35520
aggtgctcga	cggcgagcag	tcgttgagac	ccgtcgacgt	ggtgcagccg	gtgctgttcg	35580
cggtgatggt	gtcgttgggc	cggttgtggc	agtcgtacgg	ggtgactccg	ggtgcgggtg	35640
tgggtcactc	gcagggggag	atcgccgcgc	cgcacgtggc	tgggtcggtt	tcgttgggcg	35700
acgccgccag	ggtggtggcg	ttgcgcagcc	gggtgctgcg	ccgtctcggt	ggtcacggcg	35760
ggatggcgct	gttcgggctc	caccccgcac	aggccgcgca	gcggatcgcg	cgcttcgcgg	35820
gtgcgctgac	tgtegcctcg	gtcaacggtc	ccggttcggt	ggtgctggcc	ggggagaacg	35880
gcccgttgga	cgagctgatc	gccgagtgcg	aggccgaggg	cgtgaccgcc	cgtcggatcc	35940
ccgtcgacta	cgcctcacac	tccccgcagg	tggagtcgct	gcgtgaggag	ctgctcgccg	36000
cactggccgg	ggtccgtccg	gtgtcgggcg	ggatccccct	gtactcgacc	ctgaccggtc	36060
aggtcatcga	aacggcgacg	atggacgccc	actactgggt	cgccaacctc	cgggagccgg	36120
tgcgcttcca	ggacgccacc	aggcagctcg	ccgaggcggg	gttcgacgcc	ttcgtcgagg	36180
tcagcccgcga	cccgtgtgtg	acagtcgggt	tcgaggccac	cctcgaggca	gtgctgcccc	36240
ccgacgcgga	tccgtgtgtc	acaggcaccc	tgcgccgcga	acgcggcggt	ctcgcgcagt	36300
tccacaccgc	gctcgccgag	gcgtacaccc	gggggggtgga	ggtcgactgg	cgtaccgcag	36360
tgggtgaggg	acgcccggtc	gacctgccgg	tctacccggt	ccaacgacag	aacttctggc	36420
tcccgggtccc	cctgggcccg	gtccccgcga	ccggcgacga	gtggcggttac	cagctcgccct	36480
ggcaccgccg	cgacctcggg	cggtcctccc	tggccggacg	ggtcctgggt	gtgaccggag	36540
cggcagtagc	cccggcctgg	acggacgtgg	tcgcgcagcg	cctggaacag	cgcggggcga	36600
ccgtcggtgt	gtgcaccgcg	cagtcgcgcg	cccggatcgg	cgccgcactc	gacgccgtcg	36660
acggcaccgc	cctgtccact	gtggtctctc	tgctcgcgct	cgccgagggc	ggtgctgctg	36720



acgacccccag	cctggacacc	ctcgcgttgg	tccaggcgct	cggcgcagcc	gggatcgacg	36780
tccccctgtg	gctggtgacc	agggacgccg	ccgccgtgac	cgtcggagac	gacgtcgatc	36840
cggcccaggc	catggtcggg	gggctcggcc	gggtggtggg	cgtggagtcc	cccgcccggt	36900
ggggtggcct	ggtggacctg	cgcgaggccg	acgccgactc	ggcccggtcg	ctggccgcca	36960
tactggccga	cccgcgcggc	gaggagcagt	tcgcgatccg	gcccgcaggg	gtcaccgtcg	37020
cccgctctcgt	cccggcaccg	gcccgcgcgg	cggtgtaccg	gtggacgccg	cgcgggaccg	37080
tcctggtcac	cggcggcacc	ggcggcatcg	gcgcgcacct	ggcccgtggg	ctcgccgggtg	37140
cgggcgcgga	gcacctggtg	ctgctcaaca	ggcggggagc	ggaggcggcc	ggtgccgccc	37200
acctgcgtga	cgaactgggc	gcgctcggca	cgggagtcac	catcacggcc	tgcgacgtcg	37260
ccgaccgcga	ccggttggcg	gccgtcctcg	acgccgcacg	ggcgcagggg	cgggtggtca	37320
cggcgggtgtt	ccacgccgcc	gggatctccc	ggtccacagc	ggtacaggag	ctgaccgaga	37380
gcgagttcac	cgagatcacc	gacgcgaagg	tgcgggggtac	ggcgaacctg	gccgaactct	37440
gtcccagagct	ggacgccctc	gtgctgttct	cctcgaacgc	ggcgggtgtg	ggcagcccgg	37500
ggctggcctc	ctacgcggcg	ggcaacgcct	tcctcgacgc	cttcgcccggt	cgtggtcggc	37560
gcagtgggct	gccgggtcacc	tcgatcgctt	ggggtctgtg	ggccggggcag	aacatggccc	37620
gtaccgaggg	cggcgactac	ctgcgcagcc	agggcctgcg	cgccatggac	ccgcagcggg	37680
cgatcgagga	gctgcggacc	acctgggacg	ccggggaccc	gtgggtgtcg	gtgggtggacc	37740
tggaccggga	gcggttcgtc	gaactgttca	ccgccgcccg	ccgccggccc	ctcttcgacg	37800
aactcgggtg	ggtccgcgcc	ggggccgagg	agaccgggtca	ggaatcggtt	ctcgcccggc	37860
ggctggcgtc	gatgccggag	gccgaacgtc	acgagcatgt	cgcccggctg	gtccgagccc	37920
aggtggcagc	ggtgctgggc	cacggcacgc	cgacggtgat	cgagcgtgac	gtcgcttcc	37980
gtgacctggg	attcgactcc	atgaccgccg	tcgacctgcg	gaaccggctc	gcggcggtga	38040
ccgggggtccg	ggtggccacg	accatcgctt	tcgaccaccc	gacagtggac	cgcctcaccg	38100
cgcactacct	ggaacgactc	gtcgggtgagc	cggaggcgac	gaccccggct	gcggcggtcg	38160
tcccgcaggc	acccggggag	gccgacgagc	cgatcgcgat	cgtcgggatg	gcctgccgcc	38220
tcgccggtgg	agtgcgtacc	cccgaccagt	tgtgggactt	catcgctgcc	gacggcgacg	38280
cggtcaccga	gatgccgtcg	gaccggctct	gggacctcga	cgcgctgttc	gacccggacc	38340
ccgagcggca	cggcaccagc	tactcccggc	acggcgcgtt	cctggacggg	gcggccgact	38400
tcgacgcggc	gttcttcggg	atctcgcgcg	gtgaggcggt	ggcgatggat	ccgcagcagc	38460
ggcaggctct	ggagacgacg	tgggagctgt	tcgagaacgc	cggcatcgac	ccgcactccc	38520
tgcgcggtac	ggacaccggt	gtcttctctg	gcgctgcgta	ccagggggtac	ggccagaacg	38580
cgcagggtgcc	gaaggagagt	gagggttacc	tgtctaccgg	tgttctctcg	gcggtcgcct	38640
ccggtcggat	cgcgtacgtg	ttgggggttg	aggggcccgc	gatcactgtg	gacacggcgt	38700
gttcgtcgtc	gcttgtggcg	ttgcacgtgg	cggccgggtc	gctgcgatcg	ggtgactgtg	38760
ggctcgcggg	ggcgggtggg	gtgtcgggtg	tggccgggtc	ggagggtgtc	accgagttct	38820
ccaggcaggg	cgcgctggcc	cccgcaggtc	ggtgcaagcc	cttctccgac	caggccgacg	38880
ggttcggatt	cgccgagggc	gtcgtctgtg	tgtctctgca	gcgggttgtc	gtggcggtgc	38940
gggagggggc	tcgggtgttg	ggtgtgggtg	tgggttcggc	ggtgaatcag	gatggggcga	39000
gtaatgggtt	ggcggcgccg	tcgggggttg	cgcagcagcg	ggtgattcgg	cgggcgtggg	39060
gtcgtgcggg	tgtgtcgggt	ggggatgtgg	gtgtggtgga	ggcgcgatgg	acggggacgc	39120
ggttggggga	tcgggtggag	ttggggggcg	tgttggggac	gtatgggggtg	ggtcgggggtg	39180
gggtgggtcc	ggtggtggtg	ggttcgggtg	aggcgaatgt	gggtcatgtg	caggcggcgg	39240
cgggtgtggt	gggtgtgac	aagggtggtg	tgggggttgg	tcgggggttg	gtgggtccga	39300
tgggtgtgctg	gggtgggttg	tcgggggttg	tggattgggtc	gtcgggtggg	ttgggtggtg	39360
cggatggggg	gcgggggtgg	ccgggtgggtg	tggatggggg	gcgtcggggg	gggggtgtcg	39420
cgtttggggg	gtcgggggacg	aatgctcatg	tgggtgggtgc	ggaggcgccg	gggtcgggtg	39480
tggggggcga	acggccgggtg	gagggggtcg	cgcgggggtt	ggtgggggtg	gctggtggtg	39540
tgggtgccgg	ggtgctgtcg	gcaaagaccg	aaaccgccct	gaccgagctc	gcccgcagac	39600
tgcacgacgc	cgtcgacgac	accgtcgccc	tcccggcggt	ggccgcaccc	ctcgccaccg	39660
gacgcgcccc	cctgccttac	cgggccgccc	tgttgccccg	cgaccacgac	gaactgcgcg	39720
acaggctgcg	ggcgttcacc	actggttcgg	cggctcccgg	tgtggtgtcg	ggggtggcgt	39780
cgggtggtgg	tgtggtgttt	gtttttctctg	gtcagggtgg	tcagtgggtg	gggatggcgc	39840
gggggttgtt	gtcgggttcgg	gtgtttgtgg	agtccgtggt	ggagtgtgat	gcgggtggtg	39900
cgtcgggtgg	ggggttttcg	gtgttggggg	tgttggaggg	tcgggtcgggt	gcgccgtcgt	39960
tggatcgggt	ggatgtggtg	cagccgggtg	tgttcgtggt	gatggtgtcg	ttggcgcggt	40020
tgtggcggtg	gtgtgggggt	gtgcctgcgg	cgggtggtgg	tcattcgacg	ggggagatcg	40080
cggcggcggt	ggtggcgggg	gtgttgtcgg	tgggtgatgg	tgcgcgggtg	gtggcggttc	40140
gggcgcgggc	gttgcggggc	ttggccggcc	acggcggcat	ggtctccctc	gcgggtctccg	40200
ccgaacgcgc	ccgggagctg	atcgcaccct	ggtccgaccg	gatctcgggtg	gcggcggtca	40260
actccccgac	ctcgggtggtg	gtctcgggtg	acccacaggg	cctcgccgcc	ctcgtcgccc	40320
actgcgcgga	gaccgggtgag	cgggccaaga	cgctgcctgt	ggactacgcc	tccactccg	40380

cccacgtcga	acagatccgc	gacacgatcc	tcaccgacct	ggccgacgtc	acggcgcgcc	40440
gacccgacgt	cgccctctac	tccacgctgc	acggcgcccc	ggcgccggc	acggacatgg	40500
acgcccggta	ctggtacgac	aacctgcgct	caccggtgcg	cttcgacgag	gccgtcgagg	40560
ccgcccgtgc	cgacggctac	cgggtcttcg	tcgagatgag	cccacacccg	gtcctcaccg	40620
ccgcccgtgca	ggagatcgac	gacgagacgg	tggccatcgg	ctcgtgcac	cgggacaccg	40680
gcgagcggca	cctgggtcgcc	gaactcgccc	gggcccacgt	gcacggcgta	ccagtggact	40740
ggcgggcgat	cctccccgcc	acccaccccg	ttcccttgcc	gaactacccg	ttcgaggcga	40800
cccggtactg	gctcgccccg	acggcgggcg	accaggtcgc	cgaccaccgc	taccgcgtcg	40860
actggcgggc	cctggccacc	accccgggcg	agctgtccgg	cagctacctc	gtcttcggcg	40920
acgccccgga	gaccttcggc	cacagcgtcg	agaaggccgg	cgggctcctc	gtcccgggtg	40980
ccgctcccga	ccgggagtc	ctcgcggtcg	ccctggacga	ggcgggccga	cgactcgccg	41040
gtgtgctctc	cttcgcccgc	gacaccgcca	cccacctggc	ccggcaccga	ctcctcggcg	41100
aggccgacgt	cgaggcccca	ctctggctgg	tcaccagcgg	cggcgtcgca	ctcgacgacc	41160
acgaccgat	cgactgcgac	caggcaatgg	tgtgggggat	cggacgggtg	atgggtctgg	41220
agacccccga	ccggtggggc	ggcctggtgg	acgtgaccgt	cgaacccacc	gccgaggacg	41280
gggtggtctt	cgccgcccct	ctggccgccc	acgaccacga	ggaccagggtg	gcgctgcgcg	41340
acggcatccg	ccacggccga	cggctcgtcc	gcgccccgct	gaccacccga	aacgccaggt	41400
ggacaccggc	gggcacggcg	ctcgtcacgg	gcggtacggg	tgccctcggc	ggccacgtcg	41460
cgcggtacct	ggcccgggtc	ggggtgaccg	atctcgtcct	gctcagcagg	agcggccccg	41520
acgaccccg	tgccgcccga	ctggccgccc	aactggccga	cctcgggggc	gagccgagag	41580
tcgaggcgtg	cgacgtcacc	gacggggccac	gcctgcgcgc	cctggtgcag	gagctacggg	41640
aacaggaccg	gccgggtccg	atcgtcgtcc	acaccgcagg	ggtgcccga	tcccgtcccc	41700
tcgaccgat	cgacgaactg	gagtcggtca	gcgcccgcga	ggtgaccggg	gcgcggtgc	41760
tcgacgagct	ctgcccggac	gccgacacct	tcgtcctggt	ctcctcgggg	gcgggagtg	41820
ggggtagcgc	gaacctgggc	gcgtacgcgg	cagccaacgc	ctacctggac	gccctggccc	41880
accgcccggc	ccaggcgggc	cgggcccgcga	cctcgggtcg	ctggggggcg	tgggcccggc	41940
acggcatggc	caccggcgac	ctcgacgggc	tgaccggcg	cggctcgcgg	gcgatggcac	42000
cggaccgggc	gctgcgcgcc	tgcaccaggc	ggtggaccac	ccacgacacc	tgtgtgtcgg	42060
tagccgacgt	cgactgggac	cgttcgcgcg	tgggtttcac	cgccgcccgg	cccagacccc	42120
tgatcgacga	actcgtcacc	tccgcgcggc	tggccgcccc	caccgctgcg	gcggcccccg	42180
tcccggcgat	gaccgcccga	cagctactcc	agttcacgcg	ctcgcacgtg	gccgcgatcc	42240
tcggtcacca	ggacccggac	gcggtcgggt	tggaccagcc	cttcaccgag	ctgggcttcg	42300
actcgtcac	cgccgtcggc	ctgcgcaacc	agctccagca	ggccaccggg	cggacgtgc	42360
ccgcgcgcc	ggtgttccag	caccccacgg	tacgcagact	cgccgaccac	ctcgcgcagc	42420
agctcgacgt	cggcaccgcc	cgggtcgagg	cgacgggcag	cgtcctgcgg	gacggctacc	42480
ggcgggccgg	gcagaccggc	gacgtccggt	cgtacctgga	cctgctggcg	aacctgtcgg	42540
agttccggga	gcggttcacc	gacgcggcga	gcctgggcgg	acagctggaa	ctcgtcgacc	42600
tggccgacgg	atccggcccc	gtcactgtga	tctgttgccg	gggactgcg	gcgctctccg	42660
ggcgcacga	gttcgcccga	ctcgcctcgg	cgtgcgcgg	caccgtgccg	gtgcgcgcc	42720
tcgcgcaacc	cgggtacgag	gcgggtgaac	cggtgccggc	gtcgatggag	gcagtgtcgc	42780
gggtgcaggc	ggacgcggtc	ctcgcggcac	agggcgacac	gccgttcgtg	ctggtcggac	42840
actcggcggg	ggccctgatg	gcgtacgccc	tggcgaccga	gctggccgac	cggggccacc	42900
cgccacgtgg	cgtcgtgctc	ctcgacgtgt	acccaccccg	tcaccaggag	gcggtgcacg	42960
cctggctcgg	cgagctgacc	gccgccctgt	tcgaccacga	gaccgtacgg	atggacgaca	43020
cccggctcac	ggccctgggg	gcgtacgaca	ggctgaccgg	caggtggcgt	ccgagggaca	43080
ccggtctgcc	cacgctgggt	gtggccgcca	gcgagccgat	gggggagtg	ccggacgacg	43140
gttggcagtc	cacgtggccg	ttcgggcacg	acagggtcac	ggtgcccggg	gaccacttct	43200
cgatgggtgca	ggagcacgcc	gacgcgatcg	cgcggcacat	cgacgcctgg	ttgagcgggg	43260
agagggcatg	aacacgaccg	atcgcgccgt	gctgggcccga	cgactccaga	tgatccgggg	43320
actgtactgg	ggttacggca	gcaacggaga	cccgtacccg	atgctgttgt	gcgggcacga	43380
cgacgacccg	caccgctggg	accgggggct	ggcgcgatcc	ggggtccggc	gcagccgtac	43440
cgagacgtgg	gtggtgaccg	accacgccac	cgcggtgcgg	gtgctcgacg	acccgacctt	43500
cacccggggc	accggccgga	cgcggagtg	gatgcggggc	gcgggcgcc	cggcctcgac	43560
ctgggcgcag	ccgttcctgt	acgtgcacgc	cgcgtcctgg	gacgccgaac	tgcccgaacc	43620
gcaggaggtg	gaggaccggc	tgacgggtct	cctgcctgcc	ccggggaccc	gcctggacct	43680
ggtccgcgac	ctcgcctggc	cgatggcgtc	gcgggggggt	ggcgcgagcg	accccgacgt	43740
gctgcgcgcc	gcgtgggacg	cccggttcgg	cctcgacgcc	cagctcacc	cgcagcccct	43800
ggcggtgacc	gaggcggcga	tcgccgcggg	gcccggggac	ccgcaccggc	gggcgctgtt	43860
caccgcccgc	gagatgacag	ccaccgcgtt	cgtcgacgcg	gtgctggcgg	tgaccgccac	43920
ggcgggggcg	gcccagcgtc	tcgccgacga	ccccgacgtc	gccgcccgtc	tcgtcgcgga	43980
ggtgctgcgc	ctgcatccga	cggcgcacct	ggaacggcgt	accgccggca	ccgagacggt	44040

ggtgggag	cacacggtcg	cggcgggcga	cgagggtcgtc	gtgggtggteg	ccgcccga	44100
ccgtgacg	gggtcttcg	ccgacccgga	ccgcctcgac	ccggaccggg	ccgacgccga	44160
ccgggccctg	tccgccagc	gcggtcacc	cgcccggttg	gaggagctgg	tgggtgtcct	44220
gaccaccgcc	gactgcgca	gcgtcgccaa	ggcgtgccc	ggtctcaccg	ccggtggccc	44280
ggtcgtcagg	cgacgtcggt	caccggtcct	gcgagccacc	gcccactgcc	cggtcgaaact	44340
ctgaggtgcc	tgcgatgcgc	gtcgtcttct	cctccatggc	cagcaagagc	cacctgttcg	44400
gtctcgttcc	cctcgcttg	gccttccgcg	cgccgggcca	cgagggtacgg	gtcgtcgccct	44460
caccggtct	caccgacgac	atcacggcgg	ccggactgac	ggccgtaccg	gtcggcaccg	44520
acgtcgacct	tgtcgacttc	atgaccacag	ccgggtacga	catcatcgac	tacgtccgca	44580
gcctggactt	cagcgagcgg	gacccggcca	cctccacctg	ggaccacctg	ctcggcatgc	44640
agaccgtcct	caccccgacc	ttctacgccc	tgatgagccc	ggactcgctg	gtcagaggga	44700
tgatctcctt	ctgtcggtcg	tggcgacccg	actgggtcgtc	tggaccgcag	accttcgccg	44760
cgtcgatcgc	ggcgacggtg	accggcggtg	cccacgccc	actcctgtgg	ggacccgaca	44820
tcacgggtacg	ggcccggcag	aagtctcctg	ggctgctgcc	cggacagccc	gccgccacc	44880
gggaggaccc	cctcgccgag	tggctcacct	ggtctgtgga	gaggttcggc	ggccgggtgc	44940
cgaggagcgt	cgaggagctg	gtgggtcggc	agtggacgat	cgaccccgcc	ccggtcgga	45000
tgcgcctcga	caccgggctg	aggacgggtg	gcattgcgcta	cgtcgactac	aacggcccg	45060
cggtggtgcc	ggactggctg	cacgacgagc	cgaccccgcc	acgggtctgc	ctcaccttg	45120
gcattctccag	ccgggagaac	agcatcgggc	aggtctcctg	cgacgacctg	ttgggtgcgc	45180
tcgggtgacgt	cgacgccgag	atcatcgcca	cagtggacga	gcagcagctc	gaaggcgtcg	45240
cccacgtccc	ggccaacatc	cgtacggctg	ggttcgtccc	gatgcacgca	ctgctgccga	45300
cctgcgcggc	gacggtgcac	cacggcggtc	ccggcagctg	gcacaccgcc	gccatccacg	45360
gcgtgccgca	ggtgatectg	cccgaacggt	gggacaccgg	ggtccgcgcc	cagcggaccg	45420
aggaccagg	ggcgggcac	gccctgccgg	tgcgcgagct	gacctccgac	cagctccgcg	45480
aggcggtgcg	gcggtcctg	gacgatcccg	ccttcaccgc	cggtgcggcg	cggtatgcggg	45540
ccgacatgct	cgccgagccg	tcccccgccg	aggctcgctga	cgctctgtgcg	gggtggtcg	45600
gggaacggac	cgccgtcgga	tgagcaccga	cgccaccac	gtccggctcg	gccggtgcgc	45660
cctgctgacc	agccggctct	ggctgggtac	ggcagccctc	gccggccagg	acgacgccga	45720
cgagtagcgc	ctgctcgacc	acgcccgctc	ccggggcgctc	aactgcctcg	acaccgccga	45780
cgacgactct	gcgtcgacca	gtgcccaggt	cgccgaggag	tcgggtcgcc	ggtggttggc	45840
cggggacacc	ggtcggcggg	aggagaccgt	cctgtcggtg	acgggtgggtg	tcccaccggg	45900
cgggcaggtc	ggcggggggc	gcctctccgc	ccggcagatc	atgcctcct	gtgagggtc	45960
cctgcggcgt	ctcggtgtcg	accacgtcga	cgctcctcac	ctgccccggg	tggaccgggt	46020
ggagccgtgg	gacgaggtct	ggcaggcggt	ggacgccctc	gtggcccgcc	gaaaggctctg	46080
ttacgtcggg	tcgtcgggct	tccccggatg	gcacatcgctc	gccgcccagg	agcacgccgt	46140
ccgcgctcac	cgctcggcc	tgggtgtccca	ccagtgtcgg	tacgacctga	cgctcgccca	46200
tcccgaactg	gaggtcctgc	ccgcgcgcga	ggcgtacggg	ctcggggtct	tcgccaggcc	46260
gaccgcctc	ggcgggtctgc	tcggcggcga	cggtccgggc	gccgcagccg	cacgggcgtc	46320
gggacagccg	acggcactgc	gctcggcggg	ggaggcggtac	gaggtgttct	gcagagacct	46380
cggcgagcac	cccgcgagg	tcgcaactggc	gtgggtgtctg	tcccggccc	gtgtggcggg	46440
ggcgggtcgtc	ggtgcgcgga	cgcccggaac	gctcgactcc	gcgctccgcg	cctgcggcgt	46500
cgccctcggc	gcgacggaac	tcaccgccct	ggacgggatc	ttccccgggg	tcgccgcagc	46560
agggggcgcc	ccggaggcgt	ggctacgggtg	agagcccgc	cctgacctgc	gggaaccctg	46620
gtcgggtgcg	cgggacggcc	gccgcggctc	ccgccccggg	cagccgggtg	gggtgagccg	46680
cagcaggtcc	ggcgccaccg	actcggccac	ctccccgacg	tggtcggcga	ggtagaagtg	46740
cccgcgccgg	aaggtccggg	tacggccggg	gactaccgag	tacggcagcc	agcgttgggc	46800
gtcctccacc	gtcgtcaacg	ggtcgggtgc	accgcagagg	gtgggtgatgc	cggcccgag	46860
cggcggcccc	gcctgccagg	cgtaggagcg	cagcaccggg	tggtcggccc	gcagcaccgg	46920
cagcgacatg	tccaacagcc	cctggtcggc	caatgcggcc	tcgctgaccc	cgagcctgcg	46980
catctgctcg	acgagtcctg	cctcgtcggg	caggtcgggtg	cgccgctcgt	ggaccggggg	47040
ggcgggtctgc	ccggagacga	acaaccgcag	cggtcgcacc	cccggacgag	cctccaggcg	47100
acgggcggtc	tcgtaggcga	ccagggcggc	catgctgtga	ccgaacaggg	cgaacggaac	47160
ctcgccgacg	aggtcgcgca	gcacggccgc	gacctcgtcg	gcgatctccc	cggcggtgcc	47220
gagagcccg	tcgtcacgtc	ggtcctgccg	gcccggttac	tgcaccgccc	acacgtcgac	47280
ctccggggcc	agtgcgggg	cgaggtcgag	gtacgagtcg	gcggcggtc	ccgcgtgcg	47340
gaagcagtag	agccggggcc	ggtgtccgtc	ggcgacccg	aaccggcgca	accaggtgtt	47400
catcggtgtc	tcattcgttc	ggtcgcaccg	gcaggtgggtc	gatgccgcgc	agcaggagcg	47460
accgccgcca	gacaacctcg	tcggagggga	agcccagcga	cagcttcggg	aagcggtcga	47520
acaggggccc	cagggcgacc	tctccctcca	gcttggccag	cgggcgggcc	atgcagtagt	47580
ggatgccgtg	cccgaagggtg	aggtgtcccc	ggctgtccct	ggtgacgtcg	aaccggtcg	47640
ggtcggggaa	ctgtcccggg	tcgcgggttg	ccgccccggt	ggcgatcagg	acggtgctgt	47700



```

acgccgggat cgtcaccgcc cccgatctcca cctcggcggt ggcgaaccgg gtgggtggtct 47760
ccggtggggc ctggtagcgc aggatctcct ccaccgctcc ggcagcagt gccgggtcct 47820
tccggaccag cgcgagctgg tcgggggtggg tcagcagcag gtaggtgccg atcccgatga 47880
ggctcaccga cgcctcgaat cccgccagca gcagcaccag cgcgatggag gtgagttcgt 47940
cgcggtctgag ccggtcggcg tcgtcgtcct ggaccggat c 47981

```

&lt;210&gt; 2

&lt;211&gt; 48

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 2

```

Met Gly Asp Arg Val Asn Gly His Ala Thr Pro Glu Ser Thr Gln Ser
1          5          10          15
Ala Ile Arg Phe Leu Thr Arg His Gly Gly Pro Pro Thr Ala Thr Asp
20          25          30
Asp Val His Asp Trp Leu Ala His Arg Ala Ala Glu His Arg Leu Glu
35          40          45

```

&lt;210&gt; 3

&lt;211&gt; 377

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 3

```

Met Ala Val Gly Asp Arg Arg Arg Leu Gly Arg Glu Leu Gln Met Ala
1          5          10          15
Arg Gly Leu Tyr Trp Gly Phe Gly Ala Asn Gly Asp Leu Tyr Ser Met
20          25          30
Leu Leu Ser Gly Arg Asp Asp Asp Pro Trp Thr Trp Tyr Glu Arg Leu
35          40          45
Arg Ala Ala Gly Arg Gly Pro Tyr Ala Ser Arg Ala Gly Thr Trp Val
50          55          60
Val Gly Asp His Arg Thr Ala Ala Glu Val Leu Ala Asp Pro Gly Phe
65          70          75          80
Thr His Gly Pro Pro Asp Ala Ala Arg Trp Met Gln Val Ala His Cys
85          90          95
Pro Ala Ala Ser Trp Ala Gly Pro Phe Arg Glu Phe Tyr Ala Arg Thr
100          105          110
Glu Asp Ala Ala Ser Val Thr Val Asp Ala Asp Trp Leu Gln Gln Arg
115          120          125
Cys Ala Arg Leu Val Thr Glu Leu Gly Ser Arg Phe Asp Leu Val Asn
130          135          140
Asp Phe Ala Arg Glu Val Pro Val Leu Ala Leu Gly Thr Ala Pro Ala
145          150          155          160
Leu Lys Gly Val Asp Pro Asp Arg Leu Arg Ser Trp Thr Ser Ala Thr
165          170          175
Arg Val Cys Leu Asp Ala Gln Val Ser Pro Gln Gln Leu Ala Val Thr
180          185          190
Glu Gln Ala Leu Thr Ala Leu Asp Glu Ile Asp Ala Val Thr Gly Gly
195          200          205
Arg Asp Ala Ala Val Leu Val Gly Val Val Ala Glu Leu Ala Ala Asn
210          215          220
Thr Val Gly Asn Ala Val Leu Ala Val Thr Glu Leu Pro Glu Leu Ala
225          230          235          240
Ala Arg Leu Ala Asp Asp Pro Glu Thr Ala Thr Arg Val Val Thr Glu
245          250          255
Val Ser Arg Thr Ser Pro Gly Val His Leu Glu Arg Arg Thr Ala Ala
260          265          270
Ser Asp Arg Arg Val Gly Gly Val Asp Val Pro Thr Gly Gly Glu Val
275          280          285

```

Thr Val Val Val Ala Ala Ala Asn Arg Asp Pro Glu Val Phe Thr Asp  
 290 295 300  
 Pro Asp Arg Phe Asp Val Asp Arg Gly Gly Asp Ala Glu Ile Leu Ser  
 305 310 315 320  
 Ser Arg Pro Gly Ser Pro Arg Thr Asp Leu Asp Ala Leu Val Ala Thr  
 325 330 335  
 Leu Ala Thr Ala Ala Leu Arg Ala Ala Ala Pro Val Leu Pro Arg Leu  
 340 345 350  
 Ser Arg Ser Gly Pro Val Ile Arg Arg Arg Arg Ser Pro Val Ala Arg  
 355 360 365  
 Gly Leu Ser Arg Cys Pro Val Glu Leu  
 370 375

<210> 4  
 <211> 436  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 4  
 Met Arg Val Val Phe Ser Ser Met Ala Val Asn Ser His Leu Phe Gly  
 1 5 10 15  
 Leu Val Pro Leu Ala Ser Ala Phe Gln Ala Ala Gly His Glu Val Arg  
 20 25 30  
 Val Val Ala Ser Pro Ala Leu Thr Asp Asp Val Thr Gly Ala Gly Leu  
 35 40 45  
 Thr Ala Val Pro Val Gly Asp Asp Val Glu Leu Val Glu Trp His Ala  
 50 55 60  
 His Ala Gly Gln Asp Ile Val Glu Tyr Met Arg Thr Leu Asp Trp Val  
 65 70 75 80  
 Asp Gln Ser His Thr Thr Met Ser Trp Asp Asp Leu Leu Gly Met Gln  
 85 90 95  
 Thr Thr Phe Thr Pro Thr Phe Phe Ala Leu Met Ser Pro Asp Ser Leu  
 100 105 110  
 Ile Asp Gly Met Val Glu Phe Cys Arg Ser Trp Arg Pro Asp Trp Ile  
 115 120 125  
 Val Trp Glu Pro Leu Thr Phe Ala Ala Pro Ile Ala Ala Arg Val Thr  
 130 135 140  
 Gly Thr Pro His Ala Arg Met Leu Trp Gly Pro Asp Val Ala Thr Arg  
 145 150 155 160  
 Ala Arg Gln Ser Phe Leu Arg Leu Leu Ala His Gln Glu Val Glu His  
 165 170 175  
 Arg Glu Asp Pro Leu Ala Glu Trp Phe Asp Trp Thr Leu Arg Arg Phe  
 180 185 190  
 Gly Asp Asp Pro His Leu Ser Phe Asp Glu Glu Leu Val Leu Gly Gln  
 195 200 205  
 Trp Thr Val Asp Pro Ile Pro Glu Pro Leu Arg Ile Asp Thr Gly Val  
 210 215 220  
 Arg Thr Val Gly Met Arg Tyr Val Pro Tyr Asn Gly Pro Ser Val Val  
 225 230 235 240  
 Pro Ala Trp Leu Leu Arg Glu Pro Glu Arg Arg Arg Val Cys Leu Thr  
 245 250 255  
 Leu Gly Gly Ser Arg Glu His Gly Ile Gly Gln Val Ser Ile Gly  
 260 265 270  
 Glu Met Leu Asp Ala Ile Ala Asp Ile Asp Ala Glu Phe Val Ala Thr  
 275 280 285  
 Phe Asp Asp Gln Gln Leu Val Gly Val Gly Ser Val Pro Ala Asn Val  
 290 295 300  
 Arg Thr Ala Gly Phe Val Pro Met Asn Val Leu Leu Pro Thr Cys Ala  
 305 310 315 320  
 Ala Thr Val His His Gly Gly Thr Gly Ser Trp Leu Thr Ala Ala Ile  
 325 330 335

His Gly Val Pro Gln Ile Ile Leu Ser Asp Ala Asp Thr Glu Val His  
 340 345 350  
 Ala Lys Gln Leu Gln Asp Leu Gly Ala Gly Leu Ser Leu Pro Val Ala  
 355 360 365  
 Gly Met Thr Ala Glu His Leu Arg Gly Ala Ile Glu Arg Val Leu Asp  
 370 375 380  
 Glu Pro Ala Tyr Arg Leu Gly Ala Glu Arg Met Arg Asp Gly Met Arg  
 385 390 395 400  
 Thr Asp Pro Ser Pro Ala Gln Val Val Gly Ile Cys Gln Asp Leu Ala  
 405 410 415  
 Ala Asp Arg Ala Ala Arg Gly Arg Gln Pro Arg Arg Thr Ala Glu Pro  
 420 425 430  
 His Leu Pro Arg  
 435

&lt;210&gt; 5

&lt;211&gt; 390

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 5

Met Val Thr Ser Thr Asn Leu Asp Thr Thr Ala Arg Pro Ala Leu Asn  
 1 5 10 15  
 Ser Leu Thr Gly Met Arg Phe Val Ala Phe Leu Val Phe Phe Thr  
 20 25 30  
 His Val Leu Ser Arg Leu Ile Pro Asn Ser Tyr Val Tyr Ala Asp Gly  
 35 40 45  
 Leu Asp Ala Phe Trp Gln Thr Thr Gly Arg Val Gly Val Ser Phe Phe  
 50 55 60  
 Phe Ile Leu Ser Gly Phe Val Leu Thr Trp Ser Ala Arg Ala Ser Asp  
 65 70 75 80  
 Ser Val Trp Ser Phe Trp Arg Arg Arg Val Cys Lys Leu Phe Pro Asn  
 85 90 95  
 His Leu Val Thr Ala Phe Ala Ala Val Val Leu Phe Leu Val Thr Gly  
 100 105 110  
 Gln Ala Val Ser Gly Glu Ala Leu Ile Pro Asn Leu Leu Leu Ile His  
 115 120 125  
 Ala Trp Phe Pro Ala Leu Glu Ile Ser Phe Gly Ile Asn Pro Val Ser  
 130 135 140  
 Trp Ser Leu Ala Cys Glu Ala Phe Phe Tyr Leu Cys Phe Pro Leu Phe  
 145 150 155 160  
 Leu Phe Trp Ile Ser Gly Ile Arg Pro Glu Arg Leu Trp Ala Trp Ala  
 165 170 175  
 Ala Val Val Phe Ala Ala Ile Trp Ala Val Pro Val Val Ala Asp Leu  
 180 185 190  
 Leu Leu Pro Ser Ser Pro Pro Leu Ile Pro Gly Leu Glu Tyr Ser Ala  
 195 200 205  
 Ile Gln Asp Trp Phe Leu Tyr Thr Phe Pro Ala Thr Arg Ser Leu Glu  
 210 215 220  
 Phe Ile Leu Gly Ile Ile Leu Ala Arg Ile Leu Ile Thr Gly Arg Trp  
 225 230 235 240  
 Ile Asn Val Gly Leu Leu Pro Ala Val Leu Leu Phe Pro Val Phe Phe  
 245 250 255  
 Val Ala Ser Leu Phe Leu Pro Gly Val Tyr Ala Ile Ser Ser Ser Met  
 260 265 270  
 Met Ile Leu Pro Leu Val Leu Ile Ile Ala Ser Gly Ala Thr Ala Asp  
 275 280 285  
 Leu Gln Gln Lys Arg Thr Phe Met Arg Asn Arg Val Met Val Trp Leu  
 290 295 300  
 Gly Asp Val Ser Phe Ala Leu Tyr Met Val His Phe Leu Val Ile Val  
 305 310 315 320

Tyr Gly Ala Asp Leu Leu Gly Phe Ser Gln Thr Glu Asp Ala Pro Leu  
 325 330 335  
 Gly Leu Ala Leu Phe Met Ile Ile Pro Phe Leu Ala Val Ser Leu Val  
 340 345 350  
 Leu Ser Trp Leu Leu Tyr Arg Phe Val Glu Leu Pro Val Met Arg Asn  
 355 360 365  
 Trp Ala Arg Pro Ala Ser Ala Arg Arg Lys Pro Ala Thr Glu Pro Glu  
 370 375 380  
 Gln Thr Pro Ser Arg Arg  
 385 390

&lt;210&gt; 6

&lt;211&gt; 374

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 6

Met Thr Thr Tyr Val Trp Ser Tyr Leu Leu Glu Tyr Glu Arg Glu Arg  
 1 5 10 15  
 Ala Asp Ile Leu Asp Ala Val Gln Lys Val Phe Ala Ser Gly Ser Leu  
 20 25 30  
 Ile Leu Gly Gln Ser Val Glu Asn Phe Glu Thr Glu Tyr Ala Arg Tyr  
 35 40 45  
 His Gly Ile Ala His Cys Val Gly Val Asp Asn Gly Thr Asn Ala Val  
 50 55 60  
 Lys Leu Ala Leu Glu Ser Val Gly Val Gly Arg Asp Asp Glu Val Val  
 65 70 75 80  
 Thr Val Ser Asn Thr Ala Ala Pro Thr Val Leu Ala Ile Asp Glu Ile  
 85 90 95  
 Gly Ala Arg Pro Val Phe Val Asp Val Arg Asp Glu Asp Tyr Leu Met  
 100 105 110  
 Asp Thr Asp Leu Val Glu Ala Ala Val Thr Pro Arg Thr Lys Ala Ile  
 115 120 125  
 Val Pro Val His Leu Tyr Gly Gln Cys Val Asp Met Thr Ala Leu Arg  
 130 135 140  
 Glu Leu Ala Asp Arg Arg Gly Leu Lys Leu Val Glu Asp Cys Ala Gln  
 145 150 155 160  
 Ala His Gly Ala Arg Arg Asp Gly Arg Leu Ala Gly Thr Met Ser Asp  
 165 170 175  
 Ala Ala Ala Phe Ser Phe Tyr Pro Thr Lys Val Leu Gly Ala Tyr Gly  
 180 185 190  
 Asp Gly Gly Ala Val Val Thr Asn Asp Asp Glu Thr Ala Arg Ala Leu  
 195 200 205  
 Arg Arg Leu Arg Tyr Tyr Gly Met Glu Glu Val Tyr Tyr Val Thr Arg  
 210 215 220  
 Thr Pro Gly His Asn Ser Arg Leu Asp Glu Val Gln Ala Glu Ile Leu  
 225 230 235 240  
 Arg Arg Lys Leu Thr Arg Leu Asp Ala Tyr Val Ala Gly Arg Arg Ala  
 245 250 255  
 Val Ala Gln Arg Tyr Val Asp Gly Leu Ala Asp Leu Gln Asp Ser His  
 260 265 270  
 Gly Leu Glu Leu Pro Val Val Thr Asp Gly Asn Glu His Val Phe Tyr  
 275 280 285  
 Val Tyr Val Val Arg His Pro Arg Arg Asp Glu Ile Ile Lys Arg Leu  
 290 295 300  
 Arg Asp Gly Tyr Asp Ile Ser Leu Asn Ile Ser Tyr Pro Trp Pro Val  
 305 310 315 320  
 His Thr Met Thr Gly Phe Ala His Leu Gly Val Ala Ser Gly Ser Leu  
 325 330 335  
 Pro Val Thr Glu Arg Leu Ala Gly Glu Ile Phe Ser Leu Pro Met Tyr  
 340 345 350

Pro Ser Leu Pro His Asp Leu Gln Asp Arg Val Ile Glu Ala Val Arg  
           355                          360                          365  
 Glu Val Ile Thr Gly Leu  
           370

&lt;210&gt; 7

&lt;211&gt; 257

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 7

Met Pro Asn Ser His Ser Thr Thr Ser Ser Thr Asp Val Ala Pro Tyr  
   1                          5                          10                          15  
 Glu Arg Ala Asp Ile Tyr His Asp Phe Tyr His Gly Arg Gly Lys Gly  
           20                          25                          30  
 Tyr Arg Ala Glu Ala Asp Ala Leu Val Glu Val Ala Arg Lys His Thr  
           35                          40                          45  
 Pro Gln Ala Ala Thr Leu Leu Asp Val Ala Cys Gly Thr Gly Ser His  
           50                          55                          60  
 Leu Val Glu Leu Ala Asp Ser Phe Arg Glu Val Val Gly Val Asp Leu  
   65                          70                          75                          80  
 Ser Ala Ala Met Leu Ala Thr Ala Ala Arg Asn Asp Pro Gly Arg Glu  
           85                          90                          95  
 Leu His Gln Gly Asp Met Arg Asp Phe Ser Leu Asp Arg Arg Phe Asp  
           100                          105                          110  
 Val Val Thr Cys Met Phe Ser Ser Thr Gly Tyr Leu Val Asp Glu Ala  
           115                          120                          125  
 Glu Leu Asp Arg Ala Val Ala Asn Leu Ala Gly His Leu Ala Pro Gly  
           130                          135                          140  
 Gly Thr Leu Val Val Glu Pro Trp Trp Phe Pro Glu Thr Phe Arg Pro  
   145                          150                          155                          160  
 Gly Trp Val Gly Ala Asp Leu Val Thr Ser Gly Asp Arg Arg Ile Ser  
           165                          170                          175  
 Arg Met Ser His Thr Val Pro Ala Gly Leu Pro Asp Arg Thr Ala Ser  
           180                          185                          190  
 Arg Met Thr Ile His Tyr Thr Val Gly Ser Pro Glu Ala Gly Ile Glu  
           195                          200                          205  
 His Phe Thr Glu Val His Val Met Thr Leu Phe Ala Arg Ala Ala Tyr  
           210                          215                          220  
 Glu Gln Ala Phe Gln Arg Ala Gly Leu Ser Cys Ser Tyr Val Gly His  
   225                          230                          235                          240  
 Asp Leu Phe Ser Pro Gly Leu Phe Val Gly Val Ala Ala Glu Pro Gly  
           245                          250                          255  
 Arg

&lt;210&gt; 8

&lt;211&gt; 201

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 8

Met Arg Val Glu Glu Leu Gly Ile Glu Gly Val Phe Thr Phe Thr Pro  
   1                          5                          10                          15  
 Gln Thr Phe Ala Asp Glu Arg Gly Val Phe Gly Thr Ala Tyr Gln Glu  
           20                          25                          30  
 Asp Val Phe Val Ala Ala Leu Gly Arg Pro Leu Phe Pro Val Ala Gln  
           35                          40                          45  
 Val Ser Thr Thr Arg Ser Arg Arg Gly Val Val Arg Gly Val His Phe  
           50                          55                          60  
 Thr Thr Met Pro Gly Ser Met Ala Lys Tyr Val Tyr Cys Ala Arg Gly

65		70		75		80									
Arg	Ala	Met	Asp	Phe	Ala	Val	Asp	Ile	Arg	Pro	Gly	Ser	Pro	Thr	Phe
			85						90					95	
Gly	Arg	Ala	Glu	Pro	Val	Glu	Leu	Ser	Ala	Glu	Ser	Met	Val	Gly	Leu
		100						105					110		
Tyr	Leu	Pro	Val	Gly	Met	Gly	His	Leu	Phe	Val	Ser	Leu	Glu	Asp	Asp
		115					120					125			
Thr	Thr	Leu	Val	Tyr	Leu	Met	Ser	Ala	Gly	Tyr	Val	Pro	Asp	Lys	Glu
	130					135					140				
Arg	Ala	Val	His	Pro	Leu	Asp	Pro	Glu	Leu	Ala	Leu	Pro	Ile	Pro	Ala
145					150					155					160
Asp	Leu	Asp	Leu	Val	Met	Ser	Glu	Arg	Asp	Arg	Val	Ala	Pro	Thr	Leu
			165					170					175		
Arg	Glu	Ala	Arg	Asp	Gln	Gly	Ile	Leu	Pro	Asp	Tyr	Ala	Ala	Cys	Arg
		180					185						190		
Ala	Ala	Ala	His	Arg	Val	Val	Arg	Thr							
		195					200								

<210> 9  
 <211> 328  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 9

Met	Val	Val	Leu	Gly	Ala	Ser	Gly	Phe	Leu	Gly	Ser	Ala	Val	Thr	His
1				5					10					15	
Ala	Leu	Ala	Asp	Leu	Pro	Val	Arg	Val	Arg	Leu	Val	Ala	Arg	Arg	Glu
		20					25						30		
Val	Val	Val	Pro	Ser	Gly	Ala	Val	Ala	Asp	Tyr	Glu	Thr	His	Arg	Val
		35				40						45			
Asp	Leu	Thr	Glu	Pro	Gly	Ala	Leu	Ala	Glu	Val	Val	Ala	Asp	Ala	Arg
	50				55						60				
Ala	Val	Phe	Pro	Phe	Ala	Ala	Gln	Ile	Arg	Gly	Thr	Ser	Gly	Trp	Arg
65				70					75					80	
Ile	Ser	Glu	Asp	Asp	Val	Val	Ala	Glu	Arg	Thr	Asn	Val	Gly	Leu	Val
			85						90				95		
Arg	Asp	Leu	Ile	Ala	Val	Leu	Ser	Arg	Ser	Pro	His	Ala	Pro	Val	Val
		100					105						110		
Val	Phe	Pro	Gly	Ser	Asn	Thr	Gln	Val	Gly	Arg	Val	Thr	Ala	Gly	Arg
		115				120						125			
Val	Ile	Asp	Gly	Ser	Glu	Gln	Asp	His	Pro	Glu	Gly	Val	Tyr	Asp	Arg
	130					135					140				
Gln	Lys	His	Thr	Gly	Glu	Gln	Leu	Leu	Lys	Glu	Ala	Thr	Ala	Ala	Gly
145				150					155						160
Ala	Ile	Arg	Ala	Thr	Ser	Leu	Arg	Leu	Pro	Pro	Val	Phe	Gly	Val	Pro
			165					170					175		
Ala	Ala	Gly	Thr	Ala	Asp	Asp	Arg	Gly	Val	Val	Ser	Thr	Met	Ile	Arg
		180					185						190		
Arg	Ala	Leu	Thr	Gly	Gln	Pro	Leu	Thr	Met	Trp	His	Asp	Gly	Thr	Val
	195					200						205			
Arg	Arg	Glu	Leu	Leu	Tyr	Val	Thr	Asp	Ala	Ala	Arg	Ala	Phe	Val	Thr
	210				215						220				
Ala	Leu	Asp	His	Ala	Asp	Ala	Leu	Ala	Gly	Arg	His	Phe	Leu	Leu	Gly
225				230					235						240
Thr	Gly	Arg	Ser	Trp	Pro	Leu	Gly	Glu	Val	Phe	Gln	Ala	Val	Ser	Arg
			245					250					255		
Ser	Val	Ala	Arg	His	Thr	Gly	Glu	Asp	Pro	Val	Pro	Val	Val	Ser	Val
		260					265						270		
Pro	Pro	Pro	Ala	His	Met	Asp	Pro	Ser	Asp	Leu	Arg	Ser	Val	Glu	Val
	275					280					285				
Asp	Pro	Ala	Arg	Phe	Thr	Ala	Val	Thr	Gly	Trp	Arg	Ala	Thr	Val	Thr



290 295 300  
 Met Ala Glu Ala Val Asp Arg Thr Val Ala Ala Leu Ala Pro Arg Arg  
 305 310 315 320  
 Ala Ala Ala Pro Ser Glu Pro Ser  
 325

<210> 10  
 <211> 330  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 10  
 Met Gly Thr Thr Gly Ala Gly Ser Ala Arg Val Arg Val Gly Arg Ser  
 1 5 10 15  
 Ala Leu His Thr Ser Arg Leu Trp Leu Gly Thr Val Asn Phe Ser Gly  
 20 25 30  
 Arg Val Thr Asp Asp Asp Ala Leu Arg Leu Met Asp His Ala Leu Glu  
 35 40 45  
 Arg Gly Val Asn Cys Ile Asp Thr Ala Asp Ile Tyr Gly Trp Arg Leu  
 50 55 60  
 Tyr Lys Gly His Thr Glu Glu Leu Val Gly Arg Trp Phe Ala Gln Gly  
 65 70 75 80  
 Gly Gly Arg Arg Glu Glu Thr Val Leu Ala Thr Lys Val Gly Ser Glu  
 85 90 95  
 Met Ser Glu Arg Val Asn Asp Gly Gly Leu Ser Ala Arg His Ile Val  
 100 105 110  
 Ala Ala Cys Glu Asn Ser Leu Arg Arg Leu Gly Val Asp His Ile Asp  
 115 120 125  
 Ile Tyr Gln Thr His His Ile Asp Arg Ala Ala Pro Trp Asp Glu Val  
 130 135 140  
 Trp Gln Ala Ala Glu His Leu Val Gly Ser Gly Lys Val Gly Tyr Val  
 145 150 155 160  
 Gly Ser Ser Asn Leu Ala Gly Trp His Ile Ala Ala Ala Gln Glu Ser  
 165 170 175  
 Ala Ala Arg Arg Asn Leu Leu Gly Met Ile Ser His Gln Cys Leu Tyr  
 180 185 190  
 Asn Leu Ala Val Arg His Pro Glu Leu Asp Val Leu Pro Ala Ala Gln  
 195 200 205  
 Ala Tyr Gly Val Gly Val Phe Ala Trp Ser Pro Leu His Gly Gly Leu  
 210 215 220  
 Leu Ser Gly Val Leu Glu Lys Leu Ala Ala Gly Thr Ala Val Lys Ser  
 225 230 235 240  
 Ala Gln Gly Arg Ala Gln Val Leu Leu Pro Ala Val Arg Pro Leu Val  
 245 250 255  
 Glu Ala Tyr Glu Asp Tyr Cys Arg Arg Leu Gly Ala Asp Pro Ala Glu  
 260 265 270  
 Val Gly Leu Ala Trp Val Leu Ser Arg Pro Gly Ile Leu Gly Ala Val  
 275 280 285  
 Ile Gly Pro Arg Thr Pro Glu Gln Leu Asp Ser Ala Leu Arg Ala Ala  
 290 295 300  
 Glu Leu Thr Leu Gly Glu Glu Glu Leu Arg Glu Leu Glu Ala Ile Phe  
 305 310 315 320  
 Pro Ala Pro Ala Val Asp Gly Pro Val Pro  
 325 330

<210> 11  
 <211> 417  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 11

Met	Arg	Val	Leu	Leu	Thr	Ser	Phe	Ala	His	Arg	Thr	His	Phe	Gln	Gly	1	5	10	15
Leu	Val	Pro	Leu	Ala	Trp	Ala	Leu	His	Thr	Ala	Gly	His	Asp	Val	Arg	20	25	30	
Val	Ala	Ser	Gln	Pro	Glu	Leu	Thr	Asp	Val	Val	Val	Gly	Ala	Gly	Leu	35	40	45	
Thr	Ser	Val	Pro	Leu	Gly	Ser	Asp	His	Arg	Leu	Phe	Asp	Ile	Ser	Pro	50	55	60	
Glu	Ala	Ala	Ala	Gln	Val	His	Arg	Tyr	Thr	Thr	Asp	Leu	Asp	Phe	Ala	65	70	75	80
Arg	Arg	Gly	Pro	Glu	Leu	Arg	Ser	Trp	Glu	Phe	Leu	His	Gly	Ile	Glu	85	90	95	
Glu	Ala	Thr	Ser	Arg	Phe	Val	Phe	Pro	Val	Val	Asn	Asn	Asp	Ser	Phe	100	105	110	
Val	Asp	Glu	Leu	Val	Glu	Phe	Ala	Met	Asp	Trp	Arg	Pro	Asp	Leu	Val	115	120	125	
Leu	Trp	Glu	Pro	Phe	Thr	Phe	Ala	Gly	Ala	Val	Ala	Ala	Lys	Ala	Cys	130	135	140	
Gly	Ala	Ala	His	Ala	Arg	Leu	Leu	Trp	Gly	Ser	Asp	Leu	Thr	Gly	Tyr	145	150	155	160
Phe	Arg	Ser	Arg	Ser	Gln	Asp	Leu	Arg	Gly	Gln	Arg	Pro	Ala	Asp	Asp	165	170	175	
Arg	Pro	Asp	Pro	Leu	Gly	Gly	Trp	Leu	Thr	Glu	Val	Ala	Gly	Arg	Phe	180	185	190	
Gly	Leu	Asp	Tyr	Ser	Glu	Asp	Leu	Ala	Val	Gly	Gln	Trp	Ser	Val	Asp	195	200	205	
Gln	Leu	Pro	Glu	Ser	Phe	Arg	Leu	Glu	Thr	Gly	Leu	Glu	Ser	Val	His	210	215	220	
Thr	Arg	Thr	Leu	Pro	Tyr	Asn	Gly	Ser	Ser	Val	Val	Pro	Gln	Trp	Leu	225	230	235	240
Arg	Thr	Ser	Asp	Gly	Val	Arg	Arg	Val	Cys	Phe	Thr	Gly	Gly	Tyr	Ser	245	250	255	
Ala	Leu	Gly	Ile	Thr	Ser	Asn	Pro	Gln	Glu	Phe	Leu	Arg	Thr	Leu	Ala	260	265	270	
Thr	Leu	Ala	Arg	Phe	Asp	Gly	Glu	Ile	Val	Val	Thr	Arg	Ser	Gly	Leu	275	280	285	
Asp	Pro	Ala	Ser	Val	Pro	Asp	Asn	Val	Arg	Leu	Val	Asp	Phe	Val	Pro	290	295	300	
Met	Asn	Ile	Leu	Leu	Pro	Gly	Cys	Ala	Ala	Val	Ile	His	His	Gly	Gly	305	310	315	320
Ala	Gly	Ser	Trp	Ala	Thr	Ala	Leu	His	His	Gly	Val	Pro	Gln	Ile	Ser	325	330	335	
Val	Ala	His	Glu	Trp	Asp	Cys	Val	Leu	Arg	Gly	Gln	Arg	Thr	Ala	Glu	340	345	350	
Leu	Gly	Ala	Gly	Val	Phe	Leu	Arg	Pro	Asp	Glu	Val	Asp	Ala	Asp	Thr	355	360	365	
Leu	Trp	Gln	Ala	Leu	Ala	Thr	Val	Val	Glu	Asp	Arg	Ser	His	Ala	Glu	370	375	380	
Asn	Ala	Glu	Lys	Leu	Arg	Gln	Glu	Ala	Leu	Ala	Ala	Pro	Thr	Pro	Ala	385	390	395	400
Glu	Val	Val	Pro	Val	Leu	Glu	Ala	Leu	Ala	His	Gln	His	Arg	Ala	Asp	405	410	415	
Arg																			

&lt;210&gt; 12

&lt;211&gt; 313

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 12

Met Thr Arg His Val Thr Leu Leu Gly Val Ser Gly Phe Val Gly Ser  
 1 5 10 15  
 Ala Leu Leu Arg Glu Phe Thr Thr His Pro Leu Arg Leu Arg Ala Val  
 20 25 30  
 Ala Arg Thr Gly Ser Arg Asp Gln Pro Pro Gly Ser Ala Gly Ile Glu  
 35 40 45  
 His Leu Arg Val Asp Leu Leu Glu Pro Gly Arg Val Ala Gln Val Val  
 50 55 60  
 Ala Asp Thr Asp Val Val Val His Leu Val Ala Tyr Ala Ala Gly Gly  
 65 70 75 80  
 Ser Thr Trp Arg Ser Ala Ala Thr Val Pro Glu Ala Glu Arg Val Asn  
 85 90 95  
 Ala Gly Ile Met Arg Asp Leu Val Ala Ala Leu Arg Ala Arg Pro Gly  
 100 105 110  
 Pro Ala Pro Val Leu Leu Phe Ala Ser Thr Thr Gln Ala Ala Asn Pro  
 115 120 125  
 Ala Ala Pro Ser Arg Tyr Ala Gln His Lys Ile Glu Ala Glu Arg Ile  
 130 135 140  
 Leu Arg Gln Ala Thr Glu Asp Gly Val Val Asp Gly Val Ile Leu Arg  
 145 150 155 160  
 Leu Pro Ala Ile Tyr Gly His Ser Gly Pro Ser Gly Gln Thr Gly Arg  
 165 170 175  
 Gly Val Val Thr Ala Met Ile Arg Arg Ala Leu Ala Gly Glu Pro Ile  
 180 185 190  
 Thr Met Trp His Glu Gly Ser Val Arg Arg Asn Leu Leu His Val Glu  
 195 200 205  
 Asp Val Ala Thr Ala Phe Thr Ala Ala Leu His Asn His Glu Ala Leu  
 210 215 220  
 Val Gly Asp Val Trp Thr Pro Ser Ala Asp Glu Ala Arg Pro Leu Gly  
 225 230 235 240  
 Glu Ile Phe Glu Thr Val Ala Ala Ser Val Ala Arg Gln Thr Gly Asn  
 245 250 255  
 Pro Ala Val Pro Val Val Ser Val Pro Pro Pro Glu Asn Ala Glu Ala  
 260 265 270  
 Asn Asp Phe Arg Ser Asp Asp Phe Asp Ser Thr Glu Phe Arg Thr Leu  
 275 280 285  
 Thr Gly Trp His Pro Arg Val Pro Leu Ala Glu Gly Ile Asp Arg Thr  
 290 295 300  
 Val Ala Ala Leu Ile Ser Thr Lys Glu  
 305 310

&lt;210&gt; 13

&lt;211&gt; 3546

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 13

Met Val Asp Val Pro Asp Leu Leu Gly Thr Arg Thr Pro His Pro Gly  
 1 5 10 15  
 Pro Leu Pro Phe Pro Trp Pro Leu Cys Gly His Asn Glu Pro Glu Leu  
 20 25 30  
 Arg Ala Arg Ala Arg Gln Leu His Ala Tyr Leu Glu Gly Ile Ser Glu  
 35 40 45  
 Asp Asp Val Val Ala Val Gly Ala Ala Leu Ala Arg Glu Thr Arg Ala  
 50 55 60  
 Gln Asp Gly Pro His Arg Ala Val Val Val Ala Ser Ser Val Thr Glu  
 65 70 75 80  
 Leu Thr Ala Ala Leu Ala Ala Leu Ala Gln Gly Arg Pro His Pro Ser  
 85 90 95  
 Val Val Arg Gly Val Ala Arg Pro Thr Ala Pro Val Val Phe Val Leu  
 100 105 110

Pro	Gly	Gln	Gly	Ala	Gln	Trp	Pro	Gly	Met	Ala	Thr	Arg	Leu	Leu	Ala
		115					120					125			
Glu	Ser	Pro	Val	Phe	Ala	Ala	Ala	Met	Arg	Ala	Cys	Glu	Arg	Ala	Phe
	130					135					140				
Asp	Glu	Val	Thr	Asp	Trp	Ser	Leu	Thr	Glu	Val	Leu	Asp	Ser	Pro	Glu
145					150					155					160
His	Leu	Arg	Arg	Val	Glu	Val	Val	Gln	Pro	Ala	Leu	Phe	Ala	Val	Gln
				165					170						175
Thr	Ser	Leu	Ala	Ala	Leu	Trp	Arg	Ser	Phe	Gly	Val	Arg	Pro	Asp	Ala
			180					185					190		
Val	Leu	Gly	His	Ser	Ile	Gly	Glu	Leu	Ala	Ala	Ala	Glu	Val	Cys	Gly
		195					200					205			
Ala	Val	Asp	Val	Glu	Ala	Ala	Ala	Arg	Ala	Ala	Ala	Leu	Trp	Ser	Arg
	210					215					220				
Glu	Met	Val	Pro	Leu	Val	Gly	Arg	Gly	Asp	Met	Ala	Ala	Val	Ala	Leu
225					230				235						240
Ser	Pro	Ala	Glu	Leu	Ala	Ala	Arg	Val	Glu	Arg	Trp	Asp	Asp	Asp	Val
				245					250					255	
Val	Pro	Ala	Gly	Val	Asn	Gly	Pro	Arg	Ser	Val	Leu	Leu	Thr	Gly	Ala
			260					265						270	
Pro	Glu	Pro	Ile	Ala	Arg	Arg	Val	Ala	Glu	Leu	Ala	Ala	Gln	Gly	Val
		275					280					285			
Arg	Ala	Gln	Val	Val	Asn	Val	Ser	Met	Ala	Ala	His	Ser	Ala	Gln	Val
	290					295					300				
Asp	Ala	Val	Ala	Glu	Gly	Met	Arg	Ser	Ala	Leu	Thr	Trp	Phe	Ala	Pro
305					310					315					320
Gly	Asp	Ser	Asp	Val	Pro	Tyr	Tyr	Ala	Gly	Leu	Thr	Gly	Gly	Arg	Leu
				325					330					335	
Asp	Thr	Arg	Glu	Leu	Gly	Ala	Asp	His	Trp	Pro	Arg	Ser	Phe	Arg	Leu
			340				345						350		
Pro	Val	Arg	Phe	Asp	Glu	Ala	Thr	Arg	Ala	Val	Leu	Glu	Leu	Gln	Pro
		355					360					365			
Gly	Thr	Phe	Ile	Glu	Ser	Ser	Pro	His	Pro	Val	Leu	Ala	Ala	Ser	Leu
	370					375					380				
Gln	Gln	Thr	Leu	Asp	Glu	Val	Gly	Ser	Pro	Ala	Ala	Ile	Val	Pro	Thr
385					390					395					400
Leu	Gln	Arg	Asp	Gln	Gly	Gly	Leu	Arg	Arg	Phe	Leu	Leu	Ala	Val	Ala
				405					410					415	
Gln	Ala	Tyr	Thr	Gly	Gly	Val	Thr	Val	Asp	Trp	Thr	Ala	Ala	Tyr	Pro
			420					425					430		
Gly	Val	Thr	Pro	Gly	His	Leu	Pro	Ser	Ala	Val	Ala	Val	Glu	Thr	Asp
		435					440					445			
Glu	Gly	Pro	Ser	Thr	Glu	Phe	Asp	Trp	Ala	Ala	Pro	Asp	His	Val	Leu
	450					455				460					
Arg	Ala	Arg	Leu	Leu	Glu	Ile	Val	Gly	Ala	Glu	Thr	Ala	Ala	Leu	Ala
465					470					475					480
Gly	Arg	Glu	Val	Asp	Ala	Arg	Ala	Thr	Phe	Arg	Glu	Leu	Gly	Leu	Asp
				485					490					495	
Ser	Val	Leu	Ala	Val	Gln	Leu	Arg	Thr	Arg	Leu	Ala	Thr	Ala	Thr	Gly
			500					505					510		
Arg	Asp	Leu	His	Ile	Ala	Met	Leu	Tyr	Asp	His	Pro	Thr	Pro	His	Ala
		515					520					525			
Leu	Thr	Glu	Ala	Leu	Leu	Arg	Gly	Pro	Gln	Glu	Glu	Pro	Gly	Arg	Gly
	530					535					540				
Glu	Glu	Thr	Ala	His	Pro	Thr	Glu	Ala	Glu	Pro	Asp	Glu	Pro	Val	Ala
545					550					555					560
Val	Val	Ala	Met	Ala	Cys	Arg	Leu	Pro	Gly	Gly	Val	Thr	Ser	Pro	Glu
				565					570					575	
Glu	Phe	Trp	Glu	Leu	Leu	Ala	Glu	Gly	Arg	Asp	Ala	Val	Gly	Gly	Leu
			580					585					590		
Pro	Thr	Asp	Arg	Gly	Trp	Asp	Leu	Asp	Ser	Leu	Phe	His	Pro	Asp	Pro

595	600	605
Thr Arg Ser Gly Thr Ala His Gln Arg Ala Gly Gly Phe Leu Thr Gly		
610	615	620
Ala Thr Ser Phe Asp Ala Ala Phe Phe Gly Leu Ser Pro Arg Glu Ala		
625	630	635
Leu Ala Val Glu Pro Gln Gln Arg Ile Thr Leu Glu Leu Ser Trp Glu		
645	650	655
Val Leu Glu Arg Ala Gly Ile Pro Pro Thr Ser Leu Arg Thr Ser Arg		
660	665	670
Thr Gly Val Phe Val Gly Leu Ile Pro Gln Glu Tyr Gly Pro Arg Leu		
675	680	685
Ala Glu Gly Gly Glu Gly Val Glu Gly Tyr Leu Met Thr Gly Thr Thr		
690	695	700
Thr Ser Val Ala Ser Gly Arg Val Ala Tyr Thr Leu Gly Leu Glu Gly		
705	710	715
Pro Ala Ile Ser Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Val		
725	730	735
His Leu Ala Cys Gln Ser Leu Arg Arg Gly Glu Ser Thr Met Ala Leu		
740	745	750
Ala Gly Gly Val Thr Val Met Pro Thr Pro Gly Met Leu Val Asp Phe		
755	760	765
Ser Arg Met Asn Ser Leu Ala Pro Asp Gly Arg Ser Lys Ala Phe Ser		
770	775	780
Ala Ala Ala Asp Gly Phe Gly Met Ala Glu Gly Ala Gly Met Leu Leu		
785	790	795
Leu Glu Arg Leu Ser Asp Ala Arg Arg His Gly His Pro Val Leu Ala		
805	810	815
Val Ile Arg Gly Thr Ala Val Asn Ser Asp Gly Ala Ser Asn Gly Leu		
820	825	830
Ser Ala Pro Asn Gly Arg Ala Gln Val Arg Val Ile Arg Gln Ala Leu		
835	840	845
Ala Glu Ser Gly Leu Thr Pro His Thr Val Asp Val Val Glu Thr His		
850	855	860
Gly Thr Gly Thr Arg Leu Gly Asp Pro Ile Glu Ala Arg Ala Leu Ser		
865	870	875
Asp Ala Tyr Gly Gly Asp Arg Glu His Pro Leu Arg Ile Gly Ser Val		
885	890	895
Lys Ser Asn Ile Gly His Thr Gln Ala Ala Gly Val Ala Gly Leu		
900	905	910
Ile Lys Leu Val Leu Ala Met Gln Ala Gly Val Leu Pro Arg Thr Leu		
915	920	925
His Ala Asp Glu Pro Ser Pro Glu Ile Asp Trp Ser Ser Gly Ala Ile		
930	935	940
Ser Leu Leu Gln Glu Pro Ala Ala Trp Pro Ala Gly Glu Arg Pro Arg		
945	950	955
Arg Ala Gly Val Ser Ser Phe Gly Ile Ser Gly Thr Asn Ala His Ala		
965	970	975
Ile Ile Glu Glu Ala Pro Pro Thr Gly Asp Asp Thr Arg Pro Asp Arg		
980	985	990
Met Gly Pro Val Val Pro Trp Val Leu Ser Ala Ser Thr Gly Glu Ala		
995	1000	1005
Leu Arg Ala Arg Ala Ala Arg Leu Ala Gly His Leu Arg Glu His Pro		
1010	1015	1020
Asp Gln Asp Leu Asp Asp Val Ala Tyr Ser Leu Ala Thr Gly Arg Ala		
1025	1030	1035
Ala Leu Ala Tyr Arg Ser Gly Phe Val Pro Ala Asp Ala Ser Thr Ala		
1045	1050	1055
Leu Arg Ile Leu Asp Glu Leu Ala Ala Gly Gly Ser Gly Asp Ala Val		
1060	1065	1070
Thr Gly Thr Ala Arg Ala Pro Gln Arg Val Val Phe Val Phe Pro Gly		
1075	1080	1085



Gln Gly Trp Gln Trp Ala Gly Met Ala Val Asp Leu Leu Asp Gly Asp  
 1090 1095 1100  
 Pro Val Phe Ala Ser Val Leu Arg Glu Cys Ala Asp Ala Leu Glu Pro  
 1105 1110 1115 1120  
 Tyr Leu Asp Phe Glu Ile Val Pro Phe Leu Arg Ala Glu Ala Gln Arg  
 1125 1130 1135  
 Arg Thr Pro Asp His Thr Leu Ser Thr Asp Arg Val Asp Val Val Gln  
 1140 1145 1150  
 Pro Val Leu Phe Ala Val Met Val Ser Leu Ala Ala Arg Trp Arg Ala  
 1155 1160 1165  
 Tyr Gly Val Glu Pro Ala Ala Val Ile Gly His Ser Gln Gly Glu Ile  
 1170 1175 1180  
 Ala Ala Ala Cys Val Ala Gly Ala Leu Ser Leu Asp Asp Ala Ala Arg  
 1185 1190 1195 1200  
 Ala Val Ala Leu Arg Ser Arg Val Ile Ala Thr Met Pro Gly Asn Gly  
 1205 1210 1215  
 Ala Met Ala Ser Ile Ala Ala Ser Val Asp Glu Val Ala Ala Arg Ile  
 1220 1225 1230  
 Asp Gly Arg Val Glu Ile Ala Ala Val Asn Gly Pro Arg Ala Val Val  
 1235 1240 1245  
 Val Ser Gly Asp Arg Asp Asp Leu Asp Arg Leu Val Ala Ser Cys Thr  
 1250 1255 1260  
 Val Glu Gly Val Arg Ala Lys Arg Leu Pro Val Asp Tyr Ala Ser His  
 1265 1270 1275 1280  
 Ser Ser His Val Glu Ala Val Arg Asp Ala Leu His Ala Glu Leu Gly  
 1285 1290 1295  
 Glu Phe Arg Pro Leu Pro Gly Phe Val Pro Phe Tyr Ser Thr Val Thr  
 1300 1305 1310  
 Gly Arg Trp Val Glu Pro Ala Glu Leu Asp Ala Gly Tyr Trp Phe Arg  
 1315 1320 1325  
 Asn Leu Arg His Arg Val Arg Phe Ala Asp Ala Val Arg Ser Leu Ala  
 1330 1335 1340  
 Asp Gln Gly Tyr Thr Thr Phe Leu Glu Val Ser Ala His Pro Val Leu  
 1345 1350 1355 1360  
 Thr Thr Ala Ile Glu Glu Ile Gly Glu Asp Arg Gly Gly Asp Leu Val  
 1365 1370 1375  
 Ala Val His Ser Leu Arg Arg Gly Ala Gly Gly Pro Val Asp Phe Gly  
 1380 1385 1390  
 Ser Ala Leu Ala Arg Ala Phe Val Ala Gly Val Ala Val Asp Trp Glu  
 1395 1400 1405  
 Ser Ala Tyr Gln Gly Ala Gly Ala Arg Arg Val Pro Leu Pro Thr Tyr  
 1410 1415 1420  
 Pro Phe Gln Arg Glu Arg Phe Trp Leu Glu Pro Asn Pro Ala Arg Arg  
 1425 1430 1435 1440  
 Val Ala Asp Ser Asp Asp Val Ser Ser Leu Arg Tyr Arg Ile Glu Trp  
 1445 1450 1455  
 His Pro Thr Asp Pro Gly Glu Pro Gly Arg Leu Asp Gly Thr Trp Leu  
 1460 1465 1470  
 Leu Ala Thr Tyr Pro Gly Arg Ala Asp Asp Arg Val Glu Ala Ala Arg  
 1475 1480 1485  
 Gln Ala Leu Glu Ser Ala Gly Ala Arg Val Glu Asp Leu Val Val Glu  
 1490 1495 1500  
 Pro Arg Thr Gly Arg Val Asp Leu Val Arg Arg Leu Asp Ala Val Gly  
 1505 1510 1515 1520  
 Pro Val Ala Gly Val Leu Cys Leu Phe Ala Val Ala Glu Pro Ala Ala  
 1525 1530 1535  
 Glu His Ser Pro Leu Ala Val Thr Ser Leu Ser Asp Thr Leu Asp Leu  
 1540 1545 1550  
 Thr Gln Ala Val Ala Gly Ser Gly Arg Glu Cys Pro Ile Trp Val Val  
 1555 1560 1565  
 Thr Glu Asn Ala Val Ala Val Gly Pro Phe Glu Arg Leu Arg Asp Pro

1570	1575	1580
Ala His Gly Ala Leu Trp	Ala Leu Gly Arg Val	Val Ala Leu Glu Asn
1585	1590	1595
Pro Ala Val Trp Gly Gly	Leu Val Asp Val	Pro Ser Gly Ser Val Ala
1605	1610	1615
Glu Leu Ser Arg His Leu	Gly Thr Thr Leu Ser	Gly Ala Gly Glu Asp
1620	1625	1630
Gln Val Ala Leu Arg Pro	Asp Gly Thr Tyr Ala	Arg Arg Trp Cys Arg
1635	1640	1645
Ala Gly Ala Gly Gly Thr	Gly Arg Trp Gln Pro	Arg Gly Thr Val Leu
1650	1655	1660
Val Thr Gly Gly Thr Gly	Gly Val Gly Arg His	Val Ala Arg Trp Leu
1665	1670	1675
Ala Arg Gln Gly Thr Pro	Cys Leu Val Leu Ala	Ser Arg Arg Gly Pro
1685	1690	1695
Asp Ala Asp Gly Val Glu	Glu Leu Leu Thr Glu	Leu Ala Asp Leu Gly
1700	1705	1710
Thr Arg Ala Thr Val Thr	Ala Cys Asp Val Thr	Asp Arg Glu Gln Leu
1715	1720	1725
Arg Ala Leu Leu Ala Thr	Val Asp Asp Glu His	Pro Leu Ser Ala Val
1730	1735	1740
Phe His Val Ala Ala Thr	Leu Asp Asp Gly Thr	Val Glu Thr Leu Thr
1745	1750	1755
Gly Asp Arg Ile Glu Arg	Ala Asn Arg Ala Lys	Val Leu Gly Ala Arg
1765	1770	1775
Asn Leu His Glu Leu Thr	Arg Asp Ala Asp Leu	Asp Ala Phe Val Leu
1780	1785	1790
Phe Ser Ser Ser Thr Ala	Ala Phe Gly Ala Pro	Gly Leu Gly Gly Tyr
1795	1800	1805
Val Pro Gly Asn Ala Tyr	Leu Asp Gly Leu Ala	Gln Gln Arg Arg Ser
1810	1815	1820
Glu Gly Leu Pro Ala Thr	Ser Val Ala Trp Gly	Thr Trp Ala Gly Ser
1825	1830	1835
Gly Met Ala Glu Gly Pro	Val Ala Asp Arg Phe	Arg Arg His Gly Val
1845	1850	1855
Met Glu Met His Pro Asp	Gln Ala Val Glu Gly	Leu Arg Val Ala Leu
1860	1865	1870
Val Gln Gly Glu Val Ala	Pro Ile Val Val Asp	Ile Arg Trp Asp Arg
1875	1880	1885
Phe Leu Leu Ala Tyr Thr	Ala Gln Arg Pro Thr	Arg Leu Phe Asp Thr
1890	1895	1900
Leu Asp Glu Ala Arg Arg	Ala Ala Pro Gly Pro	Asp Ala Gly Pro Gly
1905	1910	1915
Val Ala Ala Leu Ala Gly	Leu Pro Val Gly Glu	Arg Glu Lys Ala Val
1925	1930	1935
Leu Asp Leu Val Arg Thr	His Ala Ala Val Leu	Gly His Ala Ser
1940	1945	1950
Ala Glu Gln Val Pro Val	Asp Arg Ala Phe Ala	Glu Leu Gly Val Asp
1955	1960	1965
Ser Leu Ser Ala Leu Glu	Leu Arg Asn Arg Leu	Thr Thr Ala Thr Gly
1970	1975	1980
Val Arg Leu Ala Thr Thr	Thr Val Phe Asp His	Pro Asp Val Arg Thr
1985	1990	1995
Leu Ala Gly His Leu Ala	Ala Glu Leu Gly Gly	Gly Ser Gly Arg Glu
2005	2010	2015
Arg Pro Gly Gly Glu Ala	Pro Thr Val Ala Pro	Thr Asp Glu Pro Ile
2020	2025	2030
Ala Ile Val Gly Met Ala	Cys Arg Leu Pro Gly	Gly Val Asp Ser Pro
2035	2040	2045
Glu Gln Leu Trp Glu Leu	Ile Val Ser Gly Arg	Asp Thr Ala Ser Ala
2050	2055	2060

Ala Pro Gly Asp Arg Ser Trp Asp Pro Ala Glu Leu Met Val Ser Asp  
 2065 2070 2075 2080  
 Thr Thr Gly Thr Arg Thr Ala Phe Gly Asn Phe Met Pro Gly Ala Gly  
 2085 2090 2095  
 Glu Phe Asp Ala Ala Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala  
 2100 2105 2110  
 Met Asp Pro Gln Gln Arg His Ala Leu Glu Thr Thr Trp Glu Ala Leu  
 2115 2120 2125  
 Glu Asn Ala Gly Ile Arg Pro Glu Ser Leu Arg Gly Thr Asp Thr Gly  
 2130 2135 2140  
 Val Phe Val Gly Met Ser His Gln Gly Tyr Ala Thr Gly Arg Pro Lys  
 2145 2150 2155 2160  
 Pro Glu Asp Glu Val Asp Gly Tyr Leu Leu Thr Gly Asn Thr Ala Ser  
 2165 2170 2175  
 Val Ala Ser Gly Arg Ile Ala Tyr Val Leu Gly Leu Glu Gly Pro Ala  
 2180 2185 2190  
 Ile Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Val  
 2195 2200 2205  
 Ala Ala Gly Ser Leu Arg Ser Gly Asp Cys Gly Leu Ala Val Ala Gly  
 2210 2215 2220  
 Gly Val Ser Val Met Ala Gly Pro Glu Val Phe Arg Glu Phe Ser Arg  
 2225 2230 2235 2240  
 Gln Gly Ala Leu Ala Pro Asp Gly Arg Cys Lys Pro Phe Ser Asp Glu  
 2245 2250 2255  
 Ala Asp Gly Phe Gly Leu Gly Glu Gly Ser Ala Phe Val Val Leu Gln  
 2260 2265 2270  
 Arg Leu Ser Val Ala Val Arg Glu Gly Arg Arg Val Leu Gly Val Val  
 2275 2280 2285  
 Val Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Ala Ala  
 2290 2295 2300  
 Pro Ser Gly Val Ala Gln Gln Arg Val Ile Arg Arg Ala Trp Gly Arg  
 2305 2310 2315 2320  
 Ala Gly Val Ser Gly Gly Asp Val Gly Val Val Glu Ala His Gly Thr  
 2325 2330 2335  
 Gly Thr Arg Leu Gly Asp Pro Val Glu Leu Gly Ala Leu Leu Gly Thr  
 2340 2345 2350  
 Tyr Gly Val Gly Arg Gly Gly Val Gly Pro Val Val Val Gly Ser Val  
 2355 2360 2365  
 Lys Ala Asn Val Gly His Val Gln Ala Ala Ala Gly Val Val Gly Val  
 2370 2375 2380  
 Ile Lys Val Val Leu Gly Leu Gly Arg Gly Leu Val Gly Pro Met Val  
 2385 2390 2395 2400  
 Cys Arg Gly Gly Leu Ser Gly Leu Val Asp Trp Ser Ser Gly Gly Leu  
 2405 2410 2415  
 Val Val Ala Asp Gly Val Arg Gly Trp Pro Val Gly Val Asp Gly Val  
 2420 2425 2430  
 Arg Arg Gly Gly Val Ser Ala Phe Gly Val Ser Gly Thr Asn Ala His  
 2435 2440 2445  
 Val Val Val Ala Glu Ala Pro Gly Ser Val Val Gly Ala Glu Arg Pro  
 2450 2455 2460  
 Val Glu Gly Ser Ser Arg Gly Leu Val Gly Val Val Gly Gly Val Val  
 2465 2470 2475 2480  
 Pro Val Val Leu Ser Ala Lys Thr Glu Thr Ala Leu His Ala Gln Ala  
 2485 2490 2495  
 Arg Arg Leu Ala Asp His Leu Glu Thr His Pro Asp Val Pro Met Thr  
 2500 2505 2510  
 Asp Val Val Trp Thr Leu Thr Gln Ala Arg Gln Arg Phe Asp Arg Arg  
 2515 2520 2525  
 Ala Val Leu Leu Ala Ala Asp Arg Thr Gln Ala Val Glu Arg Leu Arg  
 2530 2535 2540  
 Gly Leu Ala Gly Gly Glu Pro Gly Thr Gly Val Val Ser Gly Val Ala

2545	2550	2555	2560
Ser Gly Gly Gly Val Val Phe Val Phe Pro Gly Gln Gly Gly Gln Trp			
	2565	2570	2575
Val Gly Met Ala Arg Gly Leu Leu Ser Val Pro Val Phe Val Glu Ser			
	2580	2585	2590
Val Val Glu Cys Asp Ala Val Val Ser Ser Val Val Gly Phe Ser Val			
	2595	2600	2605
Leu Gly Val Leu Glu Gly Arg Ser Gly Ala Pro Ser Leu Asp Arg Val			
	2610	2615	2620
Asp Val Val Gln Pro Val Leu Phe Val Val Met Val Ser Leu Ala Arg			
2625	2630	2635	2640
Leu Trp Arg Trp Cys Gly Val Val Pro Ala Ala Val Val Gly His Ser			
	2645	2650	2655
Gln Gly Glu Ile Ala Ala Ala Val Val Ala Gly Val Leu Ser Val Gly			
	2660	2665	2670
Asp Gly Ala Arg Val Val Ala Leu Arg Ala Arg Ala Leu Arg Ala Leu			
	2675	2680	2685
Ala Gly His Gly Gly Met Ala Ser Val Arg Arg Gly Arg Asp Asp Val			
	2690	2695	2700
Gln Lys Leu Leu Asp Ser Gly Pro Trp Thr Gly Lys Leu Glu Ile Ala			
2705	2710	2715	2720
Ala Val Asn Gly Pro Asp Ala Val Val Val Ser Gly Asp Pro Arg Ala			
	2725	2730	2735
Val Thr Glu Leu Val Glu His Cys Asp Gly Ile Gly Val Arg Ala Arg			
	2740	2745	2750
Thr Ile Pro Val Asp Tyr Ala Ser His Ser Ala Gln Val Glu Ser Leu			
	2755	2760	2765
Arg Glu Glu Leu Leu Ser Val Leu Ala Gly Ile Glu Gly Arg Pro Ala			
	2770	2775	2780
Thr Val Pro Phe Tyr Ser Thr Leu Thr Gly Gly Phe Val Asp Gly Thr			
2785	2790	2795	2800
Glu Leu Asp Ala Asp Tyr Trp Tyr Arg Asn Leu Arg His Pro Val Arg			
	2805	2810	2815
Phe His Ala Ala Val Glu Ala Leu Ala Ala Arg Asp Leu Thr Thr Phe			
	2820	2825	2830
Val Glu Val Ser Pro His Pro Val Leu Ser Met Ala Val Gly Glu Thr			
	2835	2840	2845
Leu Ala Asp Val Glu Ser Ala Val Thr Val Gly Thr Leu Glu Arg Asp			
	2850	2855	2860
Thr Asp Asp Val Glu Arg Phe Leu Thr Ser Leu Ala Glu Ala His Val			
2865	2870	2875	2880
His Gly Val Pro Val Asp Trp Ala Ala Val Leu Gly Ser Gly Thr Leu			
	2885	2890	2895
Val Asp Leu Pro Thr Tyr Pro Phe Gln Gly Arg Arg Phe Trp Leu His			
	2900	2905	2910
Pro Asp Arg Gly Pro Arg Asp Asp Val Ala Asp Trp Phe His Arg Val			
	2915	2920	2925
Asp Trp Thr Ala Thr Ala Thr Asp Gly Ser Ala Arg Leu Asp Gly Arg			
	2930	2935	2940
Trp Leu Val Val Val Pro Glu Gly Tyr Thr Asp Asp Gly Trp Val Val			
2945	2950	2955	2960
Glu Val Arg Ala Ala Leu Ala Ala Gly Gly Ala Glu Pro Val Val Thr			
	2965	2970	2975
Thr Val Glu Glu Val Thr Asp Arg Val Gly Asp Ser Asp Ala Val Val			
	2980	2985	2990
Ser Met Leu Gly Leu Ala Asp Asp Gly Ala Ala Glu Thr Leu Ala Leu			
	2995	3000	3005
Leu Arg Arg Leu Asp Ala Gln Ala Ser Thr Thr Pro Leu Trp Val Val			
	3010	3015	3020
Thr Val Gly Ala Val Ala Pro Ala Gly Pro Val Gln Arg Pro Glu Gln			
3025	3030	3035	3040

Ala Thr Val Trp Gly Leu Ala Leu Val Ala Ser Leu Glu Arg Gly His  
 3045 3050 3055  
 Arg Trp Thr Gly Leu Leu Asp Leu Pro Gln Thr Pro Asp Pro Gln Leu  
 3060 3065 3070  
 Arg Pro Arg Leu Val Glu Ala Leu Ala Gly Ala Glu Asp Gln Val Ala  
 3075 3080 3085  
 Val Arg Ala Asp Ala Val His Ala Arg Arg Ile Val Pro Thr Pro Val  
 3090 3095 3100  
 Thr Gly Ala Gly Pro Tyr Thr Ala Pro Gly Gly Thr Ile Leu Val Thr  
 3105 3110 3115 3120  
 Gly Gly Thr Ala Gly Leu Gly Ala Val Thr Ala Arg Trp Leu Ala Glu  
 3125 3130 3135  
 Arg Gly Ala Glu His Leu Ala Leu Val Ser Arg Arg Gly Pro Gly Thr  
 3140 3145 3150  
 Ala Gly Val Asp Glu Val Val Arg Asp Leu Thr Gly Leu Gly Val Arg  
 3155 3160 3165  
 Val Ser Val His Ser Cys Asp Val Gly Asp Arg Glu Ser Val Gly Ala  
 3170 3175 3180  
 Leu Val Gln Glu Leu Thr Ala Ala Gly Asp Val Val Arg Gly Val Val  
 3185 3190 3195 3200  
 His Ala Ala Gly Leu Pro Gln Gln Val Pro Leu Thr Asp Met Asp Pro  
 3205 3210 3215  
 Ala Asp Leu Ala Asp Val Val Ala Val Lys Val Asp Gly Ala Val His  
 3220 3225 3230  
 Leu Ala Asp Leu Cys Pro Glu Ala Glu Leu Phe Leu Leu Phe Ser Ser  
 3235 3240 3245  
 Gly Ala Gly Val Trp Gly Ser Ala Arg Gln Gly Ala Tyr Ala Ala Gly  
 3250 3255 3260  
 Asn Ala Phe Leu Asp Ala Phe Ala Arg His Arg Arg Asp Arg Gly Leu  
 3265 3270 3275 3280  
 Pro Ala Thr Ser Val Ala Trp Gly Leu Trp Ala Ala Gly Gly Met Thr  
 3285 3290 3295  
 Gly Asp Gln Glu Ala Val Ser Phe Leu Arg Glu Arg Gly Val Arg Pro  
 3300 3305 3310  
 Met Ser Val Pro Arg Ala Leu Glu Ala Leu Glu Arg Val Leu Thr Ala  
 3315 3320 3325  
 Gly Glu Thr Ala Val Val Val Ala Asp Val Asp Trp Ala Ala Phe Ala  
 3330 3335 3340  
 Glu Ser Tyr Thr Ser Ala Arg Pro Arg Pro Leu Leu His Arg Leu Val  
 3345 3350 3355 3360  
 Thr Pro Ala Ala Ala Val Gly Glu Arg Asp Glu Pro Arg Glu Gln Thr  
 3365 3370 3375  
 Leu Arg Asp Arg Leu Ala Ala Leu Pro Arg Ala Glu Arg Ser Ala Glu  
 3380 3385 3390  
 Leu Val Arg Leu Val Arg Arg Asp Ala Ala Ala Val Leu Gly Ser Asp  
 3395 3400 3405  
 Ala Lys Ala Val Pro Ala Thr Thr Pro Phe Lys Asp Leu Gly Phe Asp  
 3410 3415 3420  
 Ser Leu Ala Ala Val Arg Phe Arg Asn Arg Leu Ala Ala His Thr Gly  
 3425 3430 3435 3440  
 Leu Arg Leu Pro Ala Thr Leu Val Phe Glu His Pro Asn Ala Ala Ala  
 3445 3450 3455  
 Val Ala Asp Leu Leu His Asp Arg Leu Gly Glu Ala Gly Glu Pro Thr  
 3460 3465 3470  
 Pro Val Arg Ser Val Gly Ala Gly Leu Ala Ala Leu Glu Gln Ala Leu  
 3475 3480 3485  
 Pro Asp Ala Ser Asp Thr Glu Arg Val Glu Leu Val Glu Arg Leu Glu  
 3490 3495 3500  
 Arg Met Leu Ala Gly Leu Arg Pro Glu Ala Gly Ala Gly Ala Asp Ala  
 3505 3510 3515 3520  
 Pro Thr Ala Gly Asp Asp Leu Gly Glu Ala Gly Val Asp Glu Leu Leu



3525                      3530                      3535  
 Asp Ala Leu Glu Arg Glu Leu Asp Ala Arg  
 3540                      3545

<210> 14  
 <211> 3562  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 14  
 Met Thr Asp Asn Asp Lys Val Ala Glu Tyr Leu Arg Arg Ala Thr Leu  
 1                      5                      10                      15  
 Asp Leu Arg Ala Ala Arg Lys Arg Leu Arg Glu Leu Gln Ser Asp Pro  
 20                      25                      30  
 Ile Ala Val Val Gly Met Ala Cys Arg Leu Pro Gly Gly Val His Leu  
 35                      40                      45  
 Pro Gln His Leu Trp Asp Leu Leu Arg Gln Gly His Glu Thr Val Ser  
 50                      55                      60  
 Thr Phe Pro Thr Gly Arg Gly Trp Asp Leu Ala Gly Leu Phe His Pro  
 65                      70                      75                      80  
 Asp Pro Asp His Pro Gly Thr Ser Tyr Val Asp Arg Gly Gly Phe Leu  
 85                      90                      95  
 Asp Asp Val Ala Gly Phe Asp Ala Glu Phe Phe Gly Ile Ser Pro Arg  
 100                      105                      110  
 Glu Ala Thr Ala Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Thr Ser  
 115                      120                      125  
 Trp Glu Leu Val Glu Ser Ala Gly Ile Asp Pro His Ser Leu Arg Gly  
 130                      135                      140  
 Thr Pro Thr Gly Val Phe Leu Gly Val Ala Arg Leu Gly Tyr Gly Glu  
 145                      150                      155                      160  
 Asn Gly Thr Glu Ala Gly Asp Ala Glu Gly Tyr Ser Val Thr Gly Val  
 165                      170                      175  
 Ala Pro Ala Val Ala Ser Gly Arg Ile Ser Tyr Ala Leu Gly Leu Glu  
 180                      185                      190  
 Gly Pro Ser Ile Ser Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala  
 195                      200                      205  
 Leu His Leu Ala Val Glu Ser Leu Arg Leu Gly Glu Ser Ser Leu Ala  
 210                      215                      220  
 Val Val Gly Gly Ala Ala Val Met Ala Thr Pro Gly Val Phe Val Asp  
 225                      230                      235                      240  
 Phe Ser Arg Gln Arg Ala Leu Ala Ala Asp Gly Arg Ser Lys Ala Phe  
 245                      250                      255  
 Gly Ala Ala Ala Asp Gly Phe Gly Phe Ser Glu Gly Val Ser Leu Val  
 260                      265                      270  
 Leu Leu Glu Arg Leu Ser Glu Ala Glu Ser Asn Gly His Glu Val Leu  
 275                      280                      285  
 Ala Val Ile Arg Gly Ser Ala Leu Asn Gln Asp Gly Ala Ser Asn Gly  
 290                      295                      300  
 Leu Ala Ala Pro Asn Gly Thr Ala Gln Arg Lys Val Ile Arg Gln Ala  
 305                      310                      315                      320  
 Leu Arg Asn Cys Gly Leu Thr Pro Ala Asp Val Asp Ala Val Glu Ala  
 325                      330                      335  
 His Gly Thr Gly Thr Thr Leu Gly Asp Pro Ile Glu Ala Asn Ala Leu  
 340                      345                      350  
 Leu Asp Thr Tyr Gly Arg Asp Arg Asp Pro Asp His Pro Leu Trp Leu  
 355                      360                      365  
 Gly Ser Val Lys Ser Asn Ile Gly His Thr Gln Ala Ala Ala Gly Val  
 370                      375                      380  
 Thr Gly Leu Leu Lys Met Val Leu Ala Leu Arg His Glu Glu Leu Pro  
 385                      390                      395                      400  
 Ala Thr Leu His Val Asp Glu Pro Thr Pro His Val Asp Trp Ser Ser

				405					410					415			
Gly	Ala	Val	Arg	Leu	Ala	Thr	Arg	Gly	Arg	Pro	Trp	Arg	Arg	Gly	Asp		
			420					425					430				
Arg	Pro	Arg	Arg	Ala	Gly	Val	Ser	Ala	Phe	Gly	Ile	Ser	Gly	Thr	Asn		
		435					440					445					
Ala	His	Val	Ile	Val	Glu	Glu	Ala	Pro	Glu	Arg	Thr	Thr	Glu	Arg	Thr		
	450				455						460						
Val	Gly	Gly	Asp	Val	Gly	Pro	Val	Pro	Leu	Val	Val	Ser	Ala	Arg	Ser		
465					470				475						480		
Ala	Ala	Ala	Leu	Arg	Ala	Gln	Ala	Ala	Gln	Val	Ala	Glu	Leu	Val	Glu		
			485						490					495			
Gly	Ser	Asp	Val	Gly	Leu	Ala	Glu	Val	Gly	Arg	Ser	Leu	Ala	Val	Thr		
		500						505					510				
Arg	Ala	Arg	His	Glu	His	Arg	Ala	Ala	Val	Val	Ala	Ser	Thr	Arg	Ala		
	515					520					525						
Glu	Ala	Val	Arg	Gly	Leu	Arg	Glu	Val	Ala	Ala	Val	Glu	Pro	Arg	Gly		
	530				535				540								
Glu	Asp	Thr	Val	Thr	Gly	Val	Ala	Glu	Thr	Ser	Gly	Arg	Thr	Val	Val		
545					550				555						560		
Phe	Leu	Phe	Pro	Gly	Gln	Gly	Ser	Gln	Trp	Val	Gly	Met	Gly	Ala	Glu		
			565					570						575			
Leu	Leu	Asp	Ser	Ala	Pro	Ala	Phe	Ala	Asp	Thr	Ile	Arg	Ala	Cys	Asp		
		580						585					590				
Glu	Ala	Met	Ala	Pro	Leu	Gln	Asp	Trp	Ser	Val	Ser	Asp	Val	Leu	Arg		
	595					600						605					
Gln	Glu	Pro	Gly	Ala	Pro	Gly	Leu	Asp	Arg	Val	Asp	Val	Val	Gln	Pro		
	610				615				620								
Val	Leu	Phe	Ala	Val	Met	Val	Ser	Leu	Ala	Arg	Leu	Trp	Gln	Ser	Tyr		
625					630				635						640		
Gly	Val	Thr	Pro	Ala	Ala	Val	Val	Gly	His	Ser	Gln	Gly	Glu	Ile	Ala		
			645					650						655			
Ala	Ala	His	Val	Ala	Gly	Ala	Leu	Ser	Leu	Ala	Asp	Ala	Ala	Arg	Leu		
		660						665					670				
Val	Val	Gly	Arg	Ser	Arg	Leu	Leu	Arg	Ser	Leu	Ser	Gly	Gly	Gly	Gly		
	675					680						685					
Met	Ser	Ala	Val	Ala	Leu	Gly	Glu	Ala	Glu	Val	Arg	Arg	Arg	Leu	Arg		
	690					695			700								
Ser	Trp	Glu	Asp	Arg	Ile	Ser	Val	Ala	Ala	Val	Asn	Gly	Pro	Arg	Ser		
705					710				715						720		
Val	Val	Val	Ala	Gly	Glu	Pro	Glu	Ala	Leu	Arg	Glu	Trp	Gly	Arg	Glu		
			725					730						735			
Arg	Glu	Ala	Glu	Gly	Val	Arg	Val	Arg	Glu	Ile	Asp	Val	Asp	Tyr	Ala		
	740							745					750				
Ser	His	Ser	Pro	Gln	Ile	Asp	Arg	Val	Arg	Asp	Glu	Leu	Leu	Thr	Val		
	755					760						765					
Thr	Gly	Glu	Ile	Glu	Pro	Arg	Ser	Ala	Glu	Ile	Thr	Phe	Tyr	Ser	Thr		
	770					775					780						
Val	Asp	Val	Arg	Ala	Val	Asp	Gly	Thr	Asp	Leu	Asp	Ala	Gly	Tyr	Trp		
785					790				795						800		
Tyr	Arg	Asn	Leu	Arg	Glu	Thr	Val	Arg	Phe	Ala	Asp	Ala	Met	Thr	Arg		
			805					810						815			
Leu	Ala	Asp	Ser	Gly	Tyr	Asp	Ala	Phe	Val	Glu	Val	Ser	Pro	His	Pro		
		820						825					830				
Val	Val	Val	Ser	Ala	Val	Ala	Glu	Ala	Val	Glu	Glu	Ala	Gly	Val	Glu		
	835						840					845					
Asp	Ala	Val	Val	Val	Gly	Thr	Leu	Ser	Arg	Gly	Asp	Gly	Gly	Pro	Gly		
	850				855						860						
Ala	Phe	Leu	Arg	Ser	Ala	Ala	Thr	Ala	His	Cys	Ala	Gly	Val	Asp	Val		
865					870				875						880		
Asp	Trp	Thr	Pro	Ala	Leu	Pro	Gly	Ala	Ala	Thr	Ile	Pro	Leu	Pro	Thr		
			885					890						895			

Tyr Pro Phe Gln Arg Lys Pro Tyr Trp Leu Arg Ser Ser Ala Pro Ala  
 900 905 910  
 Pro Ala Ser His Asp Leu Ala Tyr Arg Val Ser Trp Thr Pro Ile Thr  
 915 920 925  
 Pro Pro Gly Asp Gly Val Leu Asp Gly Asp Trp Leu Val Val His Pro  
 930 935 940  
 Gly Gly Ser Thr Gly Trp Val Asp Gly Leu Ala Ala Ala Ile Thr Ala  
 945 950 955 960  
 Gly Gly Gly Arg Val Val Ala His Pro Val Asp Ser Val Thr Ser Arg  
 965 970 975  
 Thr Gly Leu Ala Glu Ala Leu Ala Arg Arg Asp Gly Thr Phe Arg Gly  
 980 985 990  
 Val Leu Ser Trp Val Ala Thr Asp Glu Arg His Val Glu Ala Gly Ala  
 995 1000 1005  
 Val Ala Leu Leu Thr Leu Ala Gln Ala Leu Gly Asp Ala Gly Ile Asp  
 1010 1015 1020  
 Ala Pro Leu Trp Cys Leu Thr Gln Glu Ala Val Arg Thr Pro Val Asp  
 1025 1030 1035 1040  
 Gly Asp Leu Ala Arg Pro Ala Gln Ala Ala Leu His Gly Phe Ala Gln  
 1045 1050 1055  
 Val Ala Arg Leu Glu Leu Ala Arg Arg Phe Gly Gly Val Leu Asp Leu  
 1060 1065 1070  
 Pro Ala Thr Val Asp Ala Ala Gly Thr Arg Leu Val Ala Val Leu  
 1075 1080 1085  
 Ala Gly Gly Gly Glu Asp Val Val Ala Val Arg Gly Asp Arg Leu Tyr  
 1090 1095 1100  
 Gly Arg Arg Leu Val Arg Ala Thr Leu Pro Pro Pro Gly Gly Gly Phe  
 1105 1110 1115 1120  
 Thr Pro His Gly Thr Val Leu Val Thr Gly Ala Ala Gly Pro Val Gly  
 1125 1130 1135  
 Gly Arg Leu Ala Arg Trp Leu Ala Glu Arg Gly Ala Thr Arg Leu Val  
 1140 1145 1150  
 Leu Pro Gly Ala His Pro Gly Glu Glu Leu Leu Thr Ala Ile Arg Ala  
 1155 1160 1165  
 Ala Gly Ala Thr Ala Val Val Cys Glu Pro Glu Ala Glu Ala Leu Arg  
 1170 1175 1180  
 Thr Ala Ile Gly Gly Glu Leu Pro Thr Ala Leu Val His Ala Glu Thr  
 1185 1190 1195 1200  
 Leu Thr Asn Phe Ala Gly Val Ala Asp Ala Asp Pro Glu Asp Phe Ala  
 1205 1210 1215  
 Ala Thr Val Ala Ala Lys Thr Ala Leu Pro Thr Val Leu Ala Glu Val  
 1220 1225 1230  
 Leu Gly Asp His Arg Leu Glu Arg Glu Val Tyr Cys Ser Ser Val Ala  
 1235 1240 1245  
 Gly Val Trp Gly Gly Val Gly Met Ala Ala Tyr Ala Ala Gly Ser Ala  
 1250 1255 1260  
 Tyr Leu Asp Ala Leu Val Glu His Arg Arg Ala Arg Gly His Ala Ser  
 1265 1270 1275 1280  
 Ala Ser Val Ala Trp Thr Pro Trp Ala Leu Pro Gly Ala Val Asp Asp  
 1285 1290 1295  
 Gly Arg Leu Arg Glu Arg Gly Leu Arg Ser Leu Asp Val Ala Asp Ala  
 1300 1305 1310  
 Leu Gly Thr Trp Glu Arg Leu Leu Arg Ala Gly Ala Val Ser Val Ala  
 1315 1320 1325  
 Val Ala Asp Val Asp Trp Ser Val Phe Thr Glu Gly Phe Ala Ala Ile  
 1330 1335 1340  
 Arg Pro Thr Pro Leu Phe Asp Glu Leu Leu Asp Arg Arg Gly Asp Pro  
 1345 1350 1355 1360  
 Asp Gly Ala Pro Val Asp Arg Pro Gly Glu Pro Ala Gly Glu Trp Gly  
 1365 1370 1375  
 Arg Arg Ile Ala Ala Leu Ser Pro Gln Glu Gln Arg Glu Thr Leu Leu

1380	1385	1390
Thr Leu Val Gly Glu Thr Val Ala Glu Val Leu Gly His Glu Thr Gly		
1395	1400	1405
Thr Glu Ile Asn Thr Arg Arg Ala Phe Ser Glu Leu Gly Leu Asp Ser		
1410	1415	1420
Leu Gly Ser Met Ala Leu Arg Gln Arg Leu Ala Ala Arg Thr Gly Leu		
1425	1430	1435
Arg Met Pro Ala Ser Leu Val Phe Asp His Pro Thr Val Thr Ala Leu		
1445	1450	1455
Ala Arg Tyr Leu Arg Arg Leu Val Val Gly Asp Ser Asp Pro Thr Pro		
1460	1465	1470
Val Arg Val Phe Gly Pro Thr Asp Glu Ala Glu Pro Val Ala Val Val		
1475	1480	1485
Gly Ile Gly Cys Arg Phe Pro Gly Gly Ile Ala Thr Pro Glu Asp Leu		
1490	1495	1500
Trp Arg Val Val Ser Glu Gly Thr Ser Ile Thr Thr Gly Phe Pro Thr		
1505	1510	1515
Asp Arg Gly Trp Asp Leu Arg Arg Leu Tyr His Pro Asp Pro Asp His		
1525	1530	1535
Pro Gly Thr Ser Tyr Val Asp Arg Gly Gly Phe Leu Asp Gly Ala Pro		
1540	1545	1550
Asp Phe Asp Pro Gly Phe Phe Gly Ile Thr Pro Arg Glu Ala Leu Ala		
1555	1560	1565
Met Asp Pro Gln Gln Arg Leu Thr Leu Glu Ile Ala Trp Glu Ala Val		
1570	1575	1580
Glu Arg Ala Gly Ile Asp Pro Glu Thr Leu Leu Gly Ser Asp Thr Gly		
1585	1590	1595
Val Phe Val Gly Met Asn Gly Gln Ser Tyr Leu Gln Leu Leu Thr Gly		
1605	1610	1615
Glu Gly Asp Arg Leu Asn Gly Tyr Gln Gly Leu Gly Asn Ser Ala Ser		
1620	1625	1630
Val Leu Ser Gly Arg Val Ala Tyr Thr Phe Gly Trp Glu Gly Pro Ala		
1635	1640	1645
Leu Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu		
1650	1655	1660
Ala Met Gln Ser Leu Arg Arg Gly Glu Cys Ser Leu Ala Leu Ala Gly		
1665	1670	1675
Gly Val Thr Val Met Ala Asp Pro Tyr Thr Phe Val Asp Phe Ser Ala		
1685	1690	1695
Gln Arg Gly Leu Ala Ala Asp Gly Arg Cys Lys Ala Phe Ser Ala Gln		
1700	1705	1710
Ala Asp Gly Phe Ala Leu Ala Glu Gly Val Ala Ala Leu Val Leu Glu		
1715	1720	1725
Pro Leu Ser Lys Ala Arg Arg Asn Gly His Gln Val Leu Ala Val Leu		
1730	1735	1740
Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Ala Ala		
1745	1750	1755
Pro Asn Gly Pro Ser Gln Glu Arg Val Ile Arg Gln Ala Leu Thr Ala		
1765	1770	1775
Ser Gly Leu Arg Pro Ala Asp Val Asp Met Val Glu Ala His Gly Thr		
1780	1785	1790
Gly Thr Glu Leu Gly Asp Pro Ile Glu Ala Gly Ala Leu Ile Ala Ala		
1795	1800	1805
Tyr Gly Arg Asp Arg Asp Arg Pro Leu Trp Leu Gly Ser Val Lys Thr		
1810	1815	1820
Asn Ile Gly His Thr Gln Ala Ala Ala Gly Ala Ala Gly Val Ile Lys		
1825	1830	1835
Ala Val Leu Ala Met Arg His Gly Val Leu Pro Arg Ser Leu His Ala		
1845	1850	1855
Asp Glu Leu Ser Pro His Ile Asp Trp Ala Asp Gly Lys Val Glu Val		
1860	1865	1870

Leu Arg Glu Ala Arg Gln Trp Pro Pro Gly Glu Arg Pro Arg Arg Ala  
 1875 1880 1885  
 Gly Val Ser Ser Phe Gly Val Ser Gly Thr Asn Ala His Val Ile Val  
 1890 1895 1900  
 Glu Glu Ala Pro Ala Glu Pro Asp Pro Glu Pro Val Pro Ala Ala Pro  
 1905 1910 1915 1920  
 Gly Gly Pro Leu Pro Phe Val Leu His Gly Arg Ser Val Gln Thr Val  
 1925 1930 1935  
 Arg Ser Gln Ala Arg Thr Leu Ala Glu His Leu Arg Thr Thr Gly His  
 1940 1945 1950  
 Arg Asp Leu Ala Asp Thr Ala Arg Thr Leu Ala Thr Gly Arg Ala Arg  
 1955 1960 1965  
 Phe Asp Val Arg Ala Ala Val Leu Gly Thr Asp Arg Glu Gly Val Cys  
 1970 1975 1980  
 Ala Ala Leu Asp Ala Leu Ala Gln Asp Arg Pro Ser Pro Asp Val Val  
 1985 1990 1995 2000  
 Ala Pro Ala Val Phe Ala Ala Arg Thr Pro Val Leu Val Phe Pro Gly  
 2005 2010 2015  
 Gln Gly Ser Gln Trp Val Gly Met Ala Arg Asp Leu Leu Asp Ser Ser  
 2020 2025 2030  
 Glu Val Phe Ala Glu Ser Met Gly Arg Cys Ala Glu Ala Leu Ser Pro  
 2035 2040 2045-  
 Tyr Thr Asp Trp Asp Leu Leu Asp Val Val Arg Gly Val Gly Asp Pro  
 2050 2055 2060  
 Asp Pro Tyr Asp Arg Val Asp Val Leu Gln Pro Val Leu Phe Ala Val  
 2065 2070 2075 2080  
 Met Val Ser Leu Ala Arg Leu Trp Gln Ser Tyr Gly Val Thr Pro Gly  
 2085 2090 2095  
 Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala His Val Ala  
 2100 2105 2110  
 Gly Ala Leu Ser Leu Ala Asp Ala Ala Arg Val Val Ala Leu Arg Ser  
 2115 2120 2125  
 Arg Val Leu Arg Glu Leu Asp Asp Gln Gly Gly Met Val Ser Val Gly  
 2130 2135 2140  
 Thr Ser Arg Ala Glu Leu Asp Ser Val Leu Arg Arg Trp Asp Gly Arg  
 2145 2150 2155 2160  
 Val Ala Val Ala Ala Val Asn Gly Pro Gly Thr Leu Val Val Ala Gly  
 2165 2170 2175  
 Pro Thr Ala Glu Leu Asp Glu Phe Leu Ala Val Ala Glu Ala Arg Glu  
 2180 2185 2190  
 Met Arg Pro Arg Arg Ile Ala Val Arg Tyr Ala Ser His Ser Pro Glu  
 2195 2200 2205  
 Val Ala Arg Val Glu Gln Arg Leu Ala Ala Glu Leu Gly Thr Val Thr  
 2210 2215 2220  
 Ala Val Gly Gly Thr Val Pro Leu Tyr Ser Thr Ala Thr Gly Asp Leu  
 2225 2230 2235 2240  
 Leu Asp Thr Thr Ala Met Asp Ala Gly Tyr Trp Tyr Arg Asn Leu Arg  
 2245 2250 2255  
 Gln Pro Val Leu Phe Glu His Ala Val Arg Ser Leu Leu Glu Arg Gly  
 2260 2265 2270  
 Phe Glu Thr Phe Ile Glu Val Ser Pro His Pro Val Leu Leu Met Ala  
 2275 2280 2285  
 Val Glu Glu Thr Ala Glu Asp Ala Glu Arg Pro Val Thr Gly Val Pro  
 2290 2295 2300  
 Thr Leu Arg Arg Asp His Asp Gly Pro Ser Glu Phe Leu Arg Asn Leu  
 2305 2310 2315 2320  
 Leu Gly Ala His Val His Gly Val Asp Val Asp Leu Arg Pro Ala Val  
 2325 2330 2335  
 Ala His Gly Arg Leu Val Asp Leu Pro Thr Tyr Pro Phe Asp Arg Gln  
 2340 2345 2350  
 Arg Leu Trp Pro Lys Pro His Arg Arg Ala Asp Thr Ser Ser Leu Gly

2355	2360	2365
Val Arg Asp Ser Thr His	Pro Leu Leu His Ala	Ala Val Asp Val Pro
2370	2375	2380
Gly His Gly Gly Ala Val	Phe Thr Gly Arg Leu Ser	Pro Asp Glu Gln
2385	2390	2395
Gln Trp Leu Thr Gln His	Val Val Gly Gly Arg Asn	Leu Val Pro Gly
2405	2410	2415
Ser Val Leu Val Asp Leu	Ala Leu Thr Ala Gly Ala	Asp Val Gly Val
2420	2425	2430
Pro Val Leu Glu Glu Leu	Val Leu Gln Gln Pro	Leu Val Leu Thr Ala
2435	2440	2445
Ala Gly Ala Leu Leu Arg	Leu Ser Val Gly Ala	Ala Asp Glu Asp Gly
2450	2455	2460
Arg Arg Pro Val Glu Ile	His Ala Ala Glu Asp	Val Ser Asp Pro Ala
2465	2470	2475
Glu Ala Arg Trp Ser Ala	Tyr Ala Thr Gly Thr	Leu Ala Val Gly Val
2485	2490	2495
Ala Gly Gly Gly Arg Asp	Gly Thr Gln Trp Pro	Pro Pro Gly Ala Thr
2500	2505	2510
Ala Leu Thr Leu Thr Asp	His Tyr Asp Thr Leu	Ala Glu Leu Gly Tyr
2515	2520	2525
Glu Tyr Gly Pro Ala Phe	Gln Ala Leu Arg Ala	Ala Trp Gln His Gly
2530	2535	2540
Asp Val Val Tyr Ala Glu	Val Ser Leu Asp Ala	Val Glu Glu Gly Tyr
2545	2550	2555
Ala Phe Asp Pro Val Leu	Leu Asp Ala Val Ala	Gln Thr Phe Gly Leu
2565	2570	2575
Thr Ser Arg Ala Pro Gly	Lys Leu Pro Phe Ala	Trp Arg Gly Val Thr
2580	2585	2590
Leu His Ala Thr Gly Ala	Thr Ala Val Arg Val	Val Ala Thr Pro Ala
2595	2600	2605
Gly Pro Asp Ala Val Ala	Leu Arg Val Thr Asp	Pro Thr Gly Gln Leu
2610	2615	2620
Val Ala Thr Val Asp Ala	Leu Val Val Arg Asp	Ala Gly Ala Asp Arg
2625	2630	2635
Asp Gln Pro Arg Gly Arg	Asp Gly Asp Leu His	Arg Leu Glu Trp Val
2645	2650	2655
Arg Leu Ala Thr Pro Asp	Pro Thr Pro Ala Ala	Val Val His Val Ala
2660	2665	2670
Ala Asp Gly Leu Asp Asp	Leu Leu Arg Ala Gly	Gly Pro Ala Pro Gln
2675	2680	2685
Ala Val Val Val Arg Tyr	Arg Pro Asp Gly Asp	Asp Pro Thr Ala Glu
2690	2695	2700
Ala Arg His Gly Val Leu	Trp Ala Ala Thr Leu	Val Arg Arg Trp Leu
2705	2710	2715
Asp Asp Asp Arg Trp Pro	Ala Thr Thr Leu Val	Val Ala Thr Ser Ala
2725	2730	2735
Gly Val Glu Val Ser Pro	Gly Asp Asp Val Pro	Arg Pro Gly Ala Ala
2740	2745	2750
Ala Val Trp Gly Val Leu	Arg Cys Ala Gln Ala	Glu Ser Pro Asp Arg
2755	2760	2765
Phe Val Leu Val Asp Gly	Asp Pro Glu Thr Pro	Pro Ala Val Pro Asp
2770	2775	2780
Asn Pro Gln Leu Ala Val	Arg Asp Gly Ala Val	Phe Val Pro Arg Leu
2785	2790	2795
Thr Pro Leu Ala Gly Pro	Val Pro Ala Val Ala	Asp Arg Ala Tyr Arg
2805	2810	2815
Leu Val Pro Gly Asn Gly	Gly Ser Ile Glu Ala	Val Ala Phe Ala Pro
2820	2825	2830
Val Pro Asp Ala Asp Arg	Pro Leu Ala Pro Glu	Glu Val Arg Val Ala
2835	2840	2845



Val Arg Ala Thr Gly Val Asn Phe Arg Asp Val Leu Leu Ala Leu Gly  
 2850 2855 2860  
 Met Tyr Pro Glu Pro Ala Glu Met Gly Thr Glu Ala Ser Gly Val Val  
 2865 2870 2875 2880  
 Thr Glu Val Gly Ser Gly Val Arg Arg Phe Thr Pro Gly Gln Ala Val  
 2885 2890 2895  
 Thr Gly Leu Phe Gln Gly Ala Phe Gly Pro Val Ala Val Ala Asp His  
 2900 2905 2910  
 Arg Leu Leu Thr Pro Val Pro Asp Gly Trp Arg Ala Val Asp Ala Ala  
 2915 2920 2925  
 Ala Val Pro Ile Ala Phe Thr Thr Ala His Tyr Ala Leu His Asp Leu  
 2930 2935 2940  
 Ala Gly Leu Gln Ala Gly Gln Ser Val Leu Val His Ala Ala Ala Gly  
 2945 2950 2955 2960  
 Gly Val Gly Met Ala Ala Val Ala Leu Ala Arg Arg Ala Gly Ala Glu  
 2965 2970 2975  
 Val Phe Ala Thr Ala Ser Pro Ala Lys His Pro Thr Leu Arg Ala Leu  
 2980 2985 2990  
 Gly Leu Asp Asp Asp His Ile Ala Ser Ser Arg Glu Ser Gly Phe Gly  
 2995 3000 3005  
 Glu Arg Phe Ala Ala Arg Thr Gly Gly Arg Gly Val Asp Val Val Leu  
 3010 3015 3020  
 Asn Ser Leu Thr Gly Asp Leu Leu Asp Glu Ser Ala Arg Leu Leu Ala  
 3025 3030 3035 3040  
 Asp Gly Gly Val Phe Val Glu Met Gly Lys Thr Asp Leu Arg Pro Ala  
 3045 3050 3055  
 Glu Gln Phe Arg Gly Arg Tyr Val Pro Phe Asp Leu Ala Glu Ala Gly  
 3060 3065 3070  
 Pro Asp Arg Leu Gly Glu Ile Leu Glu Glu Val Val Gly Leu Leu Ala  
 3075 3080 3085  
 Ala Gly Ala Leu Asp Arg Leu Pro Val Ser Val Trp Glu Leu Ser Ala  
 3090 3095 3100  
 Ala Pro Ala Ala Leu Thr His Met Ser Arg Gly Arg His Val Gly Lys  
 3105 3110 3115 3120  
 Leu Val Leu Thr Gln Pro Ala Pro Val His Pro Asp Gly Thr Val Leu  
 3125 3130 3135  
 Val Thr Gly Gly Thr Gly Thr Leu Gly Arg Leu Val Ala Arg His Leu  
 3140 3145 3150  
 Val Thr Gly His Gly Val Pro His Leu Leu Val Ala Ser Arg Arg Gly  
 3155 3160 3165  
 Pro Ala Ala Pro Gly Ala Ala Glu Leu Arg Ala Asp Val Glu Gly Leu  
 3170 3175 3180  
 Gly Ala Thr Ile Glu Ile Val Ala Cys Asp Thr Ala Asp Arg Glu Ala  
 3185 3190 3195 3200  
 Leu Ala Ala Leu Leu Asp Ser Ile Pro Ala Asp Arg Pro Leu Thr Gly  
 3205 3210 3215  
 Val Val His Thr Ala Gly Val Leu Ala Asp Gly Leu Val Thr Ser Ile  
 3220 3225 3230  
 Asp Gly Thr Ala Thr Asp Gln Val Leu Arg Ala Lys Val Asp Ala Ala  
 3235 3240 3245  
 Trp His Leu His Asp Leu Thr Arg Asp Ala Asp Leu Ser Phe Phe Val  
 3250 3255 3260  
 Leu Phe Ser Ser Ala Ala Ser Val Leu Ala Gly Pro Gly Gln Gly Val  
 3265 3270 3275 3280  
 Tyr Ala Ala Ala Asn Gly Val Leu Asn Ala Leu Ala Gly Gln Arg Arg  
 3285 3290 3295  
 Ala Leu Gly Leu Pro Ala Lys Ala Leu Gly Trp Gly Leu Trp Ala Gln  
 3300 3305 3310  
 Ala Ser Glu Met Thr Ser Gly Leu Gly Asp Arg Ile Ala Arg Thr Gly  
 3315 3320 3325  
 Val Ala Ala Leu Pro Thr Glu Arg Ala Leu Ala Leu Phe Asp Ala Ala

3330	3335	3340
Leu Arg Ser Gly Gly Glu Val Leu Phe Pro Leu Ser Val Asp Arg Ser		
3345	3350	3355
Ala Leu Arg Arg Ala Glu Tyr Val Pro Glu Val Leu Arg Gly Ala Val		3360
3365	3370	3375
Arg Ser Thr Pro Arg Ala Ala Asn Arg Ala Glu Thr Pro Gly Arg Gly		
3380	3385	3390
Leu Leu Asp Arg Leu Val Gly Ala Pro Glu Thr Asp Gln Val Ala Ala		
3395	3400	3405
Leu Ala Glu Leu Val Arg Ser His Ala Ala Val Ala Gly Tyr Asp		
3410	3415	3420
Ser Ala Asp Gln Leu Pro Glu Arg Lys Ala Phe Lys Asp Leu Gly Phe		
3425	3430	3435
Asp Ser Leu Ala Ala Val Glu Leu Arg Asn Arg Leu Gly Val Thr Thr		3440
3445	3450	3455
Gly Val Arg Leu Pro Ser Thr Leu Val Phe Asp His Pro Thr Pro Leu		
3460	3465	3470
Ala Val Ala Glu His Leu Arg Ser Glu Leu Phe Ala Asp Ser Ala Pro		
3475	3480	3485
Asp Val Gly Val Gly Ala Arg Leu Asp Asp Leu Glu Arg Ala Leu Asp		
3490	3495	3500
Ala Leu Pro Asp Ala Gln Gly His Ala Asp Val Gly Ala Arg Leu Glu		
3505	3510	3515
Ala Leu Leu Arg Arg Trp Gln Ser Arg Arg Pro Pro Glu Thr Glu Pro		
3525	3530	3535
Val Thr Ile Ser Asp Asp Ala Ser Asp Asp Glu Leu Phe Ser Met Leu		
3540	3545	3550
Asp Arg Arg Leu Gly Gly Gly Gly Asp Val		
3555	3560	

&lt;210&gt; 15

&lt;211&gt; 3201

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 15

Met Ser Glu Ser Ser Gly Met Thr Glu Asp Arg Leu Arg Arg Tyr Leu	
1 5 10 15	
Lys Arg Thr Val Ala Glu Leu Asp Ser Val Thr Gly Arg Leu Asp Glu	
20 25 30	
Val Glu Tyr Arg Ala Arg Glu Pro Ile Ala Val Val Gly Met Ala Cys	
35 40 45	
Arg Phe Pro Gly Gly Val Asp Ser Pro Glu Ala Phe Trp Glu Phe Ile	
50 55 60	
Arg Asp Gly Gly Asp Ala Ile Ala Glu Ala Pro Thr Asp Arg Gly Trp	
65 70 75 80	
Pro Pro Ala Pro Arg Pro Arg Leu Gly Gly Leu Leu Ala Glu Pro Gly	
85 90 95	
Ala Phe Asp Ala Ala Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala	
100 105 110	
Thr Asp Pro Gln Gln Arg Leu Met Leu Glu Ile Ser Trp Glu Ala Leu	
115 120 125	
Glu Arg Ala Gly Phe Asp Pro Ser Ser Leu Arg Gly Ser Ala Gly Gly	
130 135 140	
Val Phe Thr Gly Val Gly Ala Val Asp Tyr Gly Pro Arg Pro Asp Glu	
145 150 155 160	
Ala Pro Glu Glu Val Leu Gly Tyr Val Gly Ile Gly Thr Ala Ser Ser	
165 170 175	
Val Ala Ser Gly Arg Val Ala Tyr Thr Leu Gly Leu Glu Gly Pro Ala	
180 185 190	
Val Thr Val Asp Thr Ala Cys Ser Ser Gly Leu Thr Ala Val His Leu	

195	200	205
Ala Met Glu Ser Leu Arg	Arg Asp Glu Cys Thr	Leu Val Leu Ala Gly
210	215	220
Gly Val Thr Val Met Ser	Ser Pro Gly Ala Phe	Thr Glu Phe Arg Ser
225	230	235
Gln Gly Gly Leu Ala Glu	Asp Gly Arg Cys Lys	Pro Phe Ser Arg Ala
245	250	255
Ala Asp Gly Phe Gly Leu	Ala Glu Gly Val Leu	Val Leu Gln
260	265	270
Arg Leu Ser Val Ala Arg	Ala Glu Gly Arg Pro	Val Leu Ala Val Leu
275	280	285
Arg Gly Ser Ala Ile Asn	Gln Asp Gly Ala Ser	Asn Gly Leu Thr Ala
290	295	300
Pro Ser Gly Pro Ala Gln	Arg Arg Val Ile Arg	Gln Ala Leu Glu Arg
305	310	315
Ala Arg Leu Arg Pro Val	Asp Val Asp Tyr Val	Glu Ala His Gly Thr
325	330	335
Gly Thr Arg Leu Gly Asp	Pro Ile Glu Ala His	Ala Leu Leu Asp Thr
340	345	350
Tyr Gly Ala Asp Arg Glu	Pro Gly Arg Pro Leu	Trp Val Gly Ser Val
355	360	365
Lys Ser Asn Ile Gly His	Thr Gln Ala Ala Ala	Gly Val Ala Gly Val
370	375	380
Met Lys Thr Val Leu Ala	Leu Arg His Arg Glu	Ile Pro Ala Thr Leu
385	390	395
His Phe Asp Glu Pro Ser	Pro His Val Asp Trp	Asp Arg Gly Ala Val
405	410	415
Ser Val Val Ser Glu Thr	Arg Pro Trp Pro Val	Gly Glu Arg Pro Arg
420	425	430
Arg Ala Gly Val Ser Ser	Phe Gly Ile Ser Gly	Thr Asn Ala His Val
435	440	445
Ile Val Glu Glu Ala Pro	Ser Pro Gln Ala Ala	Asp Leu Asp Pro Thr
450	455	460
Pro Gly Pro Ala Thr Gly	Ala Thr Pro Gly Thr	Asp Ala Ala Pro Thr
465	470	475
Ala Glu Pro Gly Ala Glu	Ala Val Ala Leu Val	Phe Ser Ala Arg Asp
485	490	495
Glu Arg Ala Leu Arg Ala	Gln Ala Ala Arg Leu	Ala Asp Arg Leu Thr
500	505	510
Asp Asp Pro Ala Pro Ser	Leu Arg Asp Thr Ala	Phe Thr Leu Val Thr
515	520	525
Arg Arg Ala Thr Trp Glu	His Arg Ala Val Val	Val Gly Gly Gly Glu
530	535	540
Glu Val Leu Ala Gly Leu	Arg Ala Val Ala Gly	Gly Arg Pro Val Asp
545	550	555
Gly Ala Val Ser Gly Arg	Ala Arg Ala Gly Arg	Arg Val Val Leu Val
565	570	575
Phe Pro Gly Gln Gly Ala	Gln Trp Gln Gly Met	Ala Arg Asp Leu Leu
580	585	590
Arg Gln Ser Pro Thr Phe	Ala Glu Ser Ile Asp	Ala Cys Glu Arg Ala
595	600	605
Leu Ala Pro His Val Asp	Trp Ser Leu Arg Glu	Val Leu Asp Gly Glu
610	615	620
Gln Ser Leu Asp Pro Val	Asp Val Val Gln Pro	Val Leu Phe Ala Val
625	630	635
Met Val Ser Leu Ala Arg	Leu Trp Gln Ser Tyr	Gly Val Thr Pro Gly
645	650	655
Ala Val Val Gly His Ser	Gln Gly Glu Ile Ala	Ala Ala His Val Ala
660	665	670
Gly Ala Leu Ser Leu Ala	Asp Ala Ala Arg Val	Val Ala Leu Arg Ser
675	680	685

Arg	Val	Leu	Arg	Arg	Leu	Gly	Gly	His	Gly	Gly	Met	Ala	Ser	Phe	Gly
690						695					700				
Leu	His	Pro	Asp	Gln	Ala	Ala	Glu	Arg	Ile	Ala	Arg	Phe	Ala	Gly	Ala
705					710					715					720
Leu	Thr	Val	Ala	Ser	Val	Asn	Gly	Pro	Arg	Ser	Val	Val	Leu	Ala	Gly
				725				730						735	
Glu	Asn	Gly	Pro	Leu	Asp	Glu	Leu	Ile	Ala	Glu	Cys	Glu	Ala	Glu	Gly
			740					745					750		
Val	Thr	Ala	Arg	Arg	Ile	Pro	Val	Asp	Tyr	Ala	Ser	His	Ser	Pro	Gln
		755					760					765			
Val	Glu	Ser	Leu	Arg	Glu	Glu	Leu	Leu	Ala	Ala	Leu	Ala	Gly	Val	Arg
770						775					780				
Pro	Val	Ser	Ala	Gly	Ile	Pro	Leu	Tyr	Ser	Thr	Leu	Thr	Gly	Gln	Val
785					790					795					800
Ile	Glu	Thr	Ala	Thr	Met	Asp	Ala	Asp	Tyr	Trp	Phe	Ala	Asn	Leu	Arg
				805					810					815	
Glu	Pro	Val	Arg	Phe	Gln	Asp	Ala	Thr	Arg	Gln	Leu	Ala	Glu	Ala	Gly
			820					825					830		
Phe	Asp	Ala	Phe	Val	Glu	Val	Ser	Pro	His	Pro	Val	Leu	Thr	Val	Gly
		835					840					845			
Val	Glu	Ala	Thr	Leu	Glu	Ala	Val	Leu	Pro	Pro	Asp	Ala	Asp	Pro	Cys
850						855					860				
Val	Thr	Gly	Thr	Leu	Arg	Arg	Glu	Arg	Gly	Gly	Leu	Ala	Gln	Phe	His
865					870					875					880
Thr	Ala	Leu	Ala	Glu	Ala	Tyr	Thr	Arg	Gly	Val	Glu	Val	Asp	Trp	Arg
				885					890					895	
Thr	Ala	Val	Gly	Glu	Gly	Arg	Pro	Val	Asp	Leu	Pro	Val	Tyr	Pro	Phe
			900				905						910		
Gln	Arg	Gln	Asn	Phe	Trp	Leu	Pro	Val	Pro	Leu	Gly	Arg	Val	Pro	Asp
		915					920					925			
Thr	Gly	Asp	Glu	Trp	Arg	Tyr	Gln	Leu	Ala	Trp	His	Pro	Val	Asp	Leu
930						935					940				
Gly	Arg	Ser	Ser	Leu	Ala	Gly	Arg	Val	Leu	Val	Val	Thr	Gly	Ala	Ala
945					950					955					960
Val	Pro	Pro	Ala	Trp	Thr	Asp	Val	Val	Arg	Asp	Gly	Leu	Glu	Gln	Arg
				965					970					975	
Gly	Ala	Thr	Val	Val	Leu	Cys	Thr	Ala	Gln	Ser	Arg	Ala	Arg	Ile	Gly
			980					985					990		
Ala	Ala	Leu	Asp	Ala	Val	Asp	Gly	Thr	Ala	Leu	Ser	Thr	Val	Val	Ser
		995					1000					1005			
Leu	Leu	Ala	Leu	Ala	Glu	Gly	Gly	Ala	Val	Asp	Asp	Pro	Ser	Leu	Asp
1010						1015					1020				
Thr	Leu	Ala	Leu	Val	Gln	Ala	Leu	Gly	Ala	Ala	Gly	Ile	Asp	Val	Pro
1025					1030					1035					1040
Leu	Trp	Leu	Val	Thr	Arg	Asp	Ala	Ala	Ala	Val	Thr	Val	Gly	Asp	Asp
				1045					1050					1055	
Val	Asp	Pro	Ala	Gln	Ala	Met	Val	Gly	Gly	Leu	Gly	Arg	Val	Val	Gly
			1060					1065					1070		
Val	Glu	Ser	Pro	Ala	Arg	Trp	Gly	Gly	Leu	Val	Asp	Leu	Arg	Glu	Ala
		1075					1080					1085			
Asp	Ala	Asp	Ser	Ala	Arg	Ser	Leu	Ala	Ala	Ile	Leu	Ala	Asp	Pro	Arg
1090						1095					1100				
Gly	Glu	Glu	Gln	Phe	Ala	Ile	Arg	Pro	Asp	Gly	Val	Thr	Val	Ala	Arg
1105					1110					1115					1120
Leu	Val	Pro	Ala	Pro	Ala	Arg	Ala	Ala	Gly	Thr	Arg	Trp	Thr	Pro	Arg
				1125					1130					1135	
Gly	Thr	Val	Leu	Val	Thr	Gly	Gly	Thr	Gly	Gly	Ile	Gly	Ala	His	Leu
			1140					1145					1150		
Ala	Arg	Trp	Leu	Ala	Gly	Ala	Gly	Ala	Glu	His	Leu	Val	Leu	Leu	Asn
		1155					1160					1165			
Arg	Arg	Gly	Ala	Glu	Ala	Ala	Gly	Ala	Ala	Asp	Leu	Arg	Asp	Glu	Leu

1170	1175	1180
Val Ala Leu Gly Thr Gly	Val Thr Ile Thr Ala Cys Asp Val Ala Asp	
1185	1190	1195
Arg Asp Arg Leu Ala Ala Val Leu Asp Ala Ala Arg Ala Gln Gly Arg		1200
	1205	1210
Val Val Thr Ala Val Phe His Ala Ala Gly Ile Ser Arg Ser Thr Ala		1215
	1220	1225
Val Gln Glu Leu Thr Glu Ser Glu Phe Thr Glu Ile Thr Asp Ala Lys		1230
	1235	1240
Val Arg Gly Thr Ala Asn Leu Ala Glu Leu Cys Pro Glu Leu Asp Ala		1245
	1250	1255
Leu Val Leu Phe Ser Ser Asn Ala Ala Val Trp Gly Ser Pro Gly Leu		1260
1265	1270	1275
Ala Ser Tyr Ala Ala Gly Asn Ala Phe Leu Asp Ala Phe Ala Arg Arg		1280
	1285	1290
Gly Arg Arg Ser Gly Leu Pro Val Thr Ser Ile Ala Trp Gly Leu Trp		1295
	1300	1305
Ala Gly Gln Asn Met Ala Gly Thr Glu Gly Gly Asp Tyr Leu Arg Ser		1310
	1315	1320
Gln Gly Leu Arg Ala Met Asp Pro Gln Arg Ala Ile Glu Glu Leu Arg		1325
	1330	1335
Thr Thr Leu Asp Ala Gly Asp Pro Trp Val Ser Val Val Asp Leu Asp		1340
1345	1350	1355
Arg Glu Arg Phe Val Glu Leu Phe Thr Ala Ala Arg Arg Arg Pro Leu		1360
	1365	1370
Phe Asp Glu Leu Gly Gly Val Arg Ala Gly Ala Glu Glu Thr Gly Gln		1375
	1380	1385
Glu Ser Asp Leu Ala Arg Arg Leu Ala Ser Met Pro Glu Ala Glu Arg		1390
	1395	1400
His Glu His Val Ala Arg Leu Val Arg Ala Glu Val Ala Ala Val Leu		1405
	1410	1415
Gly His Gly Thr Pro Thr Val Ile Glu Arg Asp Val Ala Phe Arg Asp		1420
1425	1430	1435
Leu Gly Phe Asp Ser Met Thr Ala Val Asp Leu Arg Asn Arg Leu Ala		1440
	1445	1450
Ala Val Thr Gly Val Arg Val Ala Thr Thr Ile Val Phe Asp His Pro		1455
	1460	1465
Thr Val Asp Arg Leu Thr Ala His Tyr Leu Glu Arg Leu Val Gly Glu		1470
	1475	1480
Pro Glu Ala Thr Thr Pro Ala Ala Ala Val Val Pro Gln Ala Pro Gly		1485
	1490	1495
Glu Ala Asp Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Leu Ala		1500
1505	1510	1515
Gly Gly Val Arg Thr Pro Asp Gln Leu Trp Asp Phe Ile Val Ala Asp		1520
	1525	1530
Gly Asp Ala Val Thr Glu Met Pro Ser Asp Arg Ser Trp Asp Leu Asp		1535
	1540	1545
Ala Leu Phe Asp Pro Asp Pro Glu Arg His Gly Thr Ser Tyr Ser Arg		1550
	1555	1560
His Gly Ala Phe Leu Asp Gly Ala Ala Asp Phe Asp Ala Ala Phe Phe		1565
	1570	1575
Gly Ile Ser Pro Arg Glu Ala Leu Ala Met Asp Pro Gln Gln Arg Gln		1580
1585	1590	1595
Val Leu Glu Thr Thr Trp Glu Leu Phe Glu Asn Ala Gly Ile Asp Pro		1600
	1605	1610
His Ser Leu Arg Gly Thr Asp Thr Gly Val Phe Leu Gly Ala Ala Tyr		1615
	1620	1625
Gln Gly Tyr Gly Gln Asn Ala Gln Val Pro Lys Glu Ser Glu Gly Tyr		1630
	1635	1640
Leu Leu Thr Gly Gly Ser Ser Ala Val Ala Ser Gly Arg Ile Ala Tyr		1645
	1650	1655
		1660

Val Leu Gly Leu Glu Gly Pro Ala Ile Thr Val Asp Thr Ala Cys Ser  
1665 1670 1675 1680  
Ser Ser Leu Val Ala Leu His Val Ala Ala Gly Ser Leu Arg Ser Gly  
1685 1690 1695  
Asp Cys Gly Leu Ala Val Ala Gly Gly Val Ser Val Met Ala Gly Pro  
1700 1705 1710  
Glu Val Phe Thr Glu Phe Ser Arg Gln Gly Ala Leu Ala Pro Asp Gly  
1715 1720 1725  
Arg Cys Lys Pro Phe Ser Asp Gln Ala Asp Gly Phe Gly Phe Ala Glu  
1730 1735 1740  
Gly Val Ala Val Val Leu Leu Gln Arg Leu Ser Val Ala Val Arg Glu  
1745 1750 1755 1760  
Gly Arg Arg Val Leu Gly Val Val Val Gly Ser Ala Val Asn Gln Asp  
1765 1770 1775  
Gly Ala Ser Asn Gly Leu Ala Ala Pro Ser Gly Val Ala Gln Gln Arg  
1780 1785 1790  
Val Ile Arg Arg Ala Trp Gly Arg Ala Gly Val Ser Gly Gly Asp Val  
1795 1800 1805  
Gly Val Val Glu Ala His Gly Thr Gly Thr Arg Leu Gly Asp Pro Val  
1810 1815 1820  
Glu Leu Gly Ala Leu Leu Gly Thr Tyr Gly Val Gly Arg Gly Gly Val  
1825 1830 1835 1840  
Gly Pro Val Val Val Gly Ser Val Lys Ala Asn Val Gly His Val Gln  
1845 1850 1855  
Ala Ala Ala Gly Val Val Gly Val Ile Lys Val Val Leu Gly Leu Gly  
1860 1865 1870  
Arg Gly Leu Val Gly Pro Met Val Cys Arg Gly Gly Leu Ser Gly Leu  
1875 1880 1885  
Val Asp Trp Ser Ser Gly Gly Leu Val Val Ala Asp Gly Val Arg Gly  
1890 1895 1900  
Trp Pro Val Gly Val Asp Gly Val Arg Arg Gly Gly Val Ser Ala Phe  
1905 1910 1915 1920  
Gly Val Ser Gly Thr Asn Ala His Val Val Val Ala Glu Ala Pro Gly  
1925 1930 1935  
Ser Val Val Gly Ala Glu Arg Pro Val Glu Gly Ser Ser Arg Gly Leu  
1940 1945 1950  
Val Gly Val Ala Gly Gly Val Val Pro Val Val Leu Ser Ala Lys Thr  
1955 1960 1965  
Glu Thr Ala Leu Thr Glu Leu Ala Arg Arg Leu His Asp Ala Val Asp  
1970 1975 1980  
Asp Thr Val Ala Leu Pro Ala Val Ala Ala Thr Leu Ala Thr Gly Arg  
1985 1990 1995 2000  
Ala His Leu Pro Tyr Arg Ala Ala Leu Leu Ala Arg Asp His Asp Glu  
2005 2010 2015  
Leu Arg Asp Arg Leu Arg Ala Phe Thr Thr Gly Ser Ala Ala Pro Gly  
2020 2025 2030  
Val Val Ser Gly Val Ala Ser Gly Gly Gly Val Val Phe Val Phe Pro  
2035 2040 2045  
Gly Gln Gly Gly Gln Trp Val Gly Met Ala Arg Gly Leu Leu Ser Val  
2050 2055 2060  
Pro Val Phe Val Glu Ser Val Val Glu Cys Asp Ala Val Val Ser Ser  
2065 2070 2075 2080  
Val Val Gly Phe Ser Val Leu Gly Val Leu Glu Gly Arg Ser Gly Ala  
2085 2090 2095  
Pro Ser Leu Asp Arg Val Asp Val Val Gln Pro Val Leu Phe Val Val  
2100 2105 2110  
Met Val Ser Leu Ala Arg Leu Trp Arg Trp Cys Gly Val Val Pro Ala  
2115 2120 2125  
Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala Val Val Ala  
2130 2135 2140  
Gly Val Leu Ser Val Gly Asp Gly Ala Arg Val Val Ala Leu Arg Ala



2145	2150	2155	2160
Arg Ala Leu Arg	Ala Leu Ala Gly His	Gly Gly Met Val Ser	Leu Ala
	2165	2170	2175
Val Ser Ala Glu Arg	Ala Arg Glu Leu Ile	Ala Pro Trp Ser Asp	Arg
	2180	2185	2190
Ile Ser Val Ala Ala	Val Asn Ser Pro Thr Ser	Val Val Val Ser	Gly
	2195	2200	2205
Asp Pro Gln Ala Leu	Ala Ala Leu Val Ala His	Cys Ala Glu Thr	Gly
	2210	2215	2220
Glu Arg Ala Lys Thr	Leu Pro Val Asp Tyr	Ala Ser His Ser	Ala His
	2225	2230	2235
Val Glu Gln Ile Arg	Asp Thr Ile Leu Thr Asp	Leu Ala Asp Val	Thr
	2245	2250	2255
Ala Arg Arg Pro Asp	Val Ala Leu Tyr Ser Thr	Leu His Gly Ala	Arg
	2260	2265	2270
Gly Ala Gly Thr Asp	Met Asp Ala Arg Tyr Trp	Tyr Asp Asn Leu	Arg
	2275	2280	2285
Ser Pro Val Arg Phe	Asp Glu Ala Val Glu Ala	Ala Val Ala Asp	Gly
	2290	2295	2300
Tyr Arg Val Phe Val	Glu Met Ser Pro His Pro	Val Leu Thr Ala	Ala
	2305	2310	2315
Val Gln Glu Ile Asp	Asp Glu Thr Val Ala Ile	Gly Ser Leu His	Arg
	2325	2330	2335
Asp Thr Gly Glu Arg	His Leu Val Ala Glu Leu	Ala Arg Ala His	Val
	2340	2345	2350
His Gly Val Pro Val	Asp Trp Arg Ala Ile Leu	Pro Ala Thr His	Pro
	2355	2360	2365
Val Pro Leu Pro Asn	Tyr Pro Phe Glu Ala Thr	Arg Tyr Trp Leu	Ala
	2370	2375	2380
Pro Thr Ala Ala Asp	Gln Val Ala Asp His Arg	Tyr Arg Val Asp	Trp
	2385	2390	2395
Arg Pro Leu Ala Thr	Thr Pro Ala Glu Leu Ser	Gly Ser Tyr Leu	Val
	2405	2410	2415
Phe Gly Asp Ala Pro	Glu Thr Leu Gly His Ser	Val Glu Lys Ala	Gly
	2420	2425	2430
Gly Leu Leu Val Pro	Val Ala Ala Pro Asp Arg	Glu Ser Leu Ala	Val
	2435	2440	2445
Ala Leu Asp Glu Ala	Ala Gly Arg Leu Ala Gly	Val Leu Ser Phe	Ala
	2450	2455	2460
Ala Asp Thr Ala Thr	His Leu Ala Arg His Arg	Leu Leu Gly Glu	Ala
	2465	2470	2475
Asp Val Glu Ala Pro	Leu Trp Leu Val Thr Ser	Gly Gly Val Ala	Leu
	2485	2490	2495
Asp Asp His Asp Pro	Ile Asp Cys Asp Gln Ala	Met Val Trp Gly	Ile
	2500	2505	2510
Gly Arg Val Met Gly	Leu Glu Thr Pro His Arg	Trp Gly Gly Leu	Val
	2515	2520	2525
Asp Val Thr Val Glu	Pro Thr Ala Glu Asp Gly	Val Val Phe Ala	Ala
	2530	2535	2540
Leu Leu Ala Ala Asp	Asp His Glu Asp Gln Val	Ala Leu Arg Asp	Gly
	2545	2550	2555
Ile Arg His Gly Arg	Arg Leu Val Arg Ala Pro	Leu Thr Thr Arg	Asn
	2565	2570	2575
Ala Arg Trp Thr Pro	Ala Gly Thr Ala Leu Val	Thr Gly Gly Thr	Gly
	2580	2585	2590
Ala Leu Gly Gly His	Val Ala Arg Tyr Leu Ala	Arg Ser Gly Val	Thr
	2595	2600	2605
Asp Leu Val Leu Leu	Ser Arg Ser Gly Pro Asp	Ala Pro Gly Ala	Ala
	2610	2615	2620
Glu Leu Ala Ala Glu	Leu Ala Asp Leu Gly Ala	Glu Pro Arg Val	Glu
	2625	2630	2635
			2640

Ala Cys Asp Val Thr Asp Gly Pro Arg Leu Arg Ala Leu Val Gln Glu  
 2645 2650 2655  
 Leu Arg Glu Gln Asp Arg Pro Val Arg Ile Val Val His Thr Ala Gly  
 2660 2665 2670  
 Val Pro Asp Ser Arg Pro Leu Asp Arg Ile Asp Glu Leu Glu Ser Val  
 2675 2680 2685  
 Ser Ala Ala Lys Val Thr Gly Ala Arg Leu Leu Asp Glu Leu Cys Pro  
 2690 2695 2700  
 Asp Ala Asp Thr Phe Val Leu Phe Ser Ser Gly Ala Gly Val Trp Gly  
 2705 2710 2715 2720  
 Ser Ala Asn Leu Gly Ala Tyr Ala Ala Asn Ala Tyr Leu Asp Ala  
 2725 2730 2735  
 Leu Ala His Arg Arg Arg Gln Ala Gly Arg Ala Ala Thr Ser Val Ala  
 2740 2745 2750  
 Trp Gly Ala Trp Ala Gly Asp Gly Met Ala Thr Gly Asp Leu Asp Gly  
 2755 2760 2765  
 Leu Thr Arg Arg Gly Leu Arg Ala Met Ala Pro Asp Arg Ala Leu Arg  
 2770 2775 2780  
 Ala Cys Thr Arg Arg Trp Thr Thr His Asp Thr Cys Val Ser Val Ala  
 2785 2790 2795 2800  
 Asp Val Asp Trp Asp Arg Phe Ala Val Gly Phe Thr Ala Ala Arg Pro  
 2805 2810 2815  
 Arg Pro Leu Ile Asp Glu Leu Val Thr Ser Ala Pro Val Ala Ala Pro  
 2820 2825 2830  
 Thr Ala Ala Ala Ala Pro Val Pro Ala Met Thr Ala Asp Gln Leu Leu  
 2835 2840 2845  
 Gln Phe Thr Arg Ser His Val Ala Ala Ile Leu Gly His Gln Asp Pro  
 2850 2855 2860  
 Asp Ala Val Gly Leu Asp Gln Pro Phe Thr Glu Leu Gly Phe Asp Ser  
 2865 2870 2875 2880  
 Leu Thr Ala Val Gly Leu Arg Asn Gln Leu Gln Gln Ala Thr Gly Arg  
 2885 2890 2895  
 Thr Leu Pro Ala Ala Leu Val Phe Gln His Pro Thr Val Arg Arg Leu  
 2900 2905 2910  
 Ala Asp His Leu Ala Gln Gln Leu Asp Val Gly Thr Ala Pro Val Glu  
 2915 2920 2925  
 Ala Thr Gly Ser Val Leu Arg Asp Gly Tyr Arg Arg Ala Gly Gln Thr  
 2930 2935 2940  
 Gly Asp Val Arg Ser Tyr Leu Asp Leu Leu Ala Asn Leu Ser Glu Phe  
 2945 2950 2955 2960  
 Arg Glu Arg Phe Thr Asp Ala Ala Ser Leu Gly Gly Gln Leu Glu Leu  
 2965 2970 2975  
 Val Asp Leu Ala Asp Gly Ser Gly Pro Val Thr Val Ile Cys Cys Ala  
 2980 2985 2990  
 Gly Thr Ala Ala Leu Ser Gly Pro His Glu Phe Ala Arg Leu Ala Ser  
 2995 3000 3005  
 Ala Leu Arg Gly Thr Val Pro Val Arg Ala Leu Ala Gln Pro Gly Tyr  
 3010 3015 3020  
 Glu Ala Gly Glu Pro Val Pro Ala Ser Met Glu Ala Val Leu Gly Val  
 3025 3030 3035 3040  
 Gln Ala Asp Ala Val Leu Ala Ala Gln Gly Asp Thr Pro Phe Val Leu  
 3045 3050 3055  
 Val Gly His Ser Ala Gly Ala Leu Met Ala Tyr Ala Leu Ala Thr Glu  
 3060 3065 3070  
 Leu Ala Asp Arg Gly His Pro Pro Arg Gly Val Val Leu Leu Asp Val  
 3075 3080 3085  
 Tyr Pro Pro Gly His Gln Glu Ala Val His Ala Trp Leu Gly Glu Leu  
 3090 3095 3100  
 Thr Ala Ala Leu Phe Asp His Glu Thr Val Arg Met Asp Asp Thr Arg  
 3105 3110 3115 3120  
 Leu Thr Ala Leu Gly Ala Tyr Asp Arg Leu Thr Gly Arg Trp Arg Pro

3125 3130 3135  
 Arg Asp Thr Gly Leu Pro Thr Leu Val Val Ala Ala Ser Glu Pro Met  
 3140 3145 3150  
 Gly Glu Trp Pro Asp Asp Gly Trp Gln Ser Thr Trp Pro Phe Gly His  
 3155 3160 3165  
 Asp Arg Val Thr Val Pro Gly Asp His Phe Ser Met Val Gln Glu His  
 3170 3175 3180  
 Ala Asp Ala Ile Ala Arg His Ile Asp Ala Trp Leu Ser Gly Glu Arg  
 3185 3190 3195 3200  
 Ala

<210> 16  
 <211> 358  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 16  
 Met Asn Thr Thr Asp Arg Ala Val Leu Gly Arg Arg Leu Gln Met Ile  
 1 5 10 15  
 Arg Gly Leu Tyr Trp Gly Tyr Gly Ser Asn Gly Asp Pro Tyr Pro Met  
 20 25 30  
 Leu Leu Cys Gly His Asp Asp Asp Pro His Arg Trp Tyr Arg Gly Leu  
 35 40 45  
 Gly Gly Ser Gly Val Arg Arg Ser Arg Thr Glu Thr Trp Val Val Thr  
 50 55 60  
 Asp His Ala Thr Ala Val Arg Val Leu Asp Asp Pro Thr Phe Thr Arg  
 65 70 75 80  
 Ala Thr Gly Arg Thr Pro Glu Trp Met Arg Ala Ala Gly Ala Pro Ala  
 85 90 95  
 Ser Thr Trp Ala Gln Pro Phe Arg Asp Val His Ala Ala Ser Trp Asp  
 100 105 110  
 Ala Glu Leu Pro Asp Pro Gln Glu Val Glu Asp Arg Leu Thr Gly Leu  
 115 120 125  
 Leu Pro Ala Pro Gly Thr Arg Leu Asp Leu Val Arg Asp Leu Ala Trp  
 130 135 140  
 Pro Met Ala Ser Arg Gly Val Gly Ala Asp Asp Pro Asp Val Leu Arg  
 145 150 155 160  
 Ala Ala Trp Asp Ala Arg Val Gly Leu Asp Ala Gln Leu Thr Pro Gln  
 165 170 175  
 Pro Leu Ala Val Thr Glu Ala Ala Ile Ala Ala Val Pro Gly Asp Pro  
 180 185 190  
 His Arg Arg Ala Leu Phe Thr Ala Val Glu Met Thr Ala Thr Ala Phe  
 195 200 205  
 Val Asp Ala Val Leu Ala Val Thr Ala Thr Ala Gly Ala Ala Gln Arg  
 210 215 220  
 Leu Ala Asp Asp Pro Asp Val Ala Ala Arg Leu Val Ala Glu Val Leu  
 225 230 235 240  
 Arg Leu His Pro Thr Ala His Leu Glu Arg Arg Thr Ala Gly Thr Glu  
 245 250 255  
 Thr Val Val Gly Glu His Thr Val Ala Ala Gly Asp Glu Val Val Val  
 260 265 270  
 Val Val Ala Ala Ala Asn Arg Asp Ala Gly Val Phe Ala Asp Pro Asp  
 275 280 285  
 Arg Leu Asp Pro Asp Arg Ala Asp Ala Asp Arg Ala Leu Ser Ala Gln  
 290 295 300  
 Arg Gly His Pro Gly Arg Leu Glu Glu Leu Val Val Val Leu Thr Thr  
 305 310 315 320  
 Ala Ala Leu Arg Ser Val Ala Lys Ala Leu Pro Gly Leu Thr Ala Gly  
 325 330 335  
 Gly Pro Val Val Arg Arg Arg Arg Ser Pro Val Leu Arg Ala Thr Ala

His Cys Pro Val Glu Leu  
355

345

350

<210> 17  
<211> 422  
<212> PRT  
<213> Micromonospora megalomicea

<400> 17  
Met Arg Val Val Phe Ser Ser Met Ala Ser Lys Ser His Leu Phe Gly  
1 5 10 15  
Leu Val Pro Leu Ala Trp Ala Phe Arg Ala Ala Gly His Glu Val Arg  
20 25 30  
Val Val Ala Ser Pro Ala Leu Thr Asp Asp Ile Thr Ala Ala Gly Leu  
35 40 45  
Thr Ala Val Pro Val Gly Thr Asp Val Asp Leu Val Asp Phe Met Thr  
50 55 60  
His Ala Gly Tyr Asp Ile Ile Asp Tyr Val Arg Ser Leu Asp Phe Ser  
65 70 75 80  
Glu Arg Asp Pro Ala Thr Ser Thr Trp Asp His Leu Leu Gly Met Gln  
85 90 95  
Thr Val Leu Thr Pro Thr Phe Tyr Ala Leu Met Ser Pro Asp Ser Leu  
100 105 110  
Val Glu Gly Met Ile Ser Phe Cys Arg Ser Trp Arg Pro Asp Trp Ser  
115 120 125  
Ser Gly Pro Gln Thr Phe Ala Ala Ser Ile Ala Ala Thr Val Thr Gly  
130 135 140  
Val Ala His Ala Arg Leu Leu Trp Gly Pro Asp Ile Thr Val Arg Ala  
145 150 155 160  
Arg Gln Lys Phe Leu Gly Leu Leu Pro Gly Gln Pro Ala Ala His Arg  
165 170 175  
Glu Asp Pro Leu Ala Glu Trp Leu Thr Trp Ser Val Glu Arg Phe Gly  
180 185 190  
Gly Arg Val Pro Gln Asp Val Glu Glu Leu Val Val Gly Gln Trp Thr  
195 200 205  
Ile Asp Pro Ala Pro Val Gly Met Arg Leu Asp Thr Gly Leu Arg Thr  
210 215 220  
Val Gly Met Arg Tyr Val Asp Tyr Asn Gly Pro Ser Val Val Pro Asp  
225 230 235 240  
Trp Leu His Asp Glu Pro Thr Arg Arg Arg Val Cys Leu Thr Leu Gly  
245 250 255  
Ile Ser Ser Arg Glu Asn Ser Ile Gly Gln Val Ser Val Asp Asp Leu  
260 265 270  
Leu Gly Ala Leu Gly Asp Val Asp Ala Glu Ile Ile Ala Thr Val Asp  
275 280 285  
Glu Gln Gln Leu Glu Gly Val Ala His Val Pro Ala Asn Ile Arg Thr  
290 295 300  
Val Gly Phe Val Pro Met His Ala Leu Leu Pro Thr Cys Ala Ala Thr  
305 310 315 320  
Val His His Gly Gly Pro Gly Ser Trp His Thr Ala Ala Ile His Gly  
325 330 335  
Val Pro Gln Val Ile Leu Pro Asp Gly Trp Asp Thr Gly Val Arg Ala  
340 345 350  
Gln Arg Thr Glu Asp Gln Gly Ala Gly Ile Ala Leu Pro Val Pro Glu  
355 360 365  
Leu Thr Ser Asp Gln Leu Arg Glu Ala Val Arg Arg Val Leu Asp Asp  
370 375 380  
Pro Ala Phe Thr Ala Gly Ala Ala Arg Met Arg Ala Asp Met Leu Ala  
385 390 395 400  
Glu Pro Ser Pro Ala Glu Val Val Asp Val Cys Ala Gly Leu Val Gly

405  
Glu Arg Thr Ala Val Gly  
420

410

415

<210> 18  
<211> 323  
<212> PRT  
<213> Micromonospora megalomicea

<400> 18  
Met Ser Thr Asp Ala Thr His Val Arg Leu Gly Arg Cys Ala Leu Leu  
1 5 10 15  
Thr Ser Arg Leu Trp Leu Gly Thr Ala Ala Leu Ala Gly Gln Asp Asp  
20 25 30  
Ala Asp Ala Val Arg Leu Leu Asp His Ala Arg Ser Arg Gly Val Asn  
35 40 45  
Cys Leu Asp Thr Ala Asp Asp Asp Ser Ala Ser Thr Ser Ala Gln Val  
50 55 60  
Ala Glu Glu Ser Val Gly Arg Trp Leu Ala Gly Asp Thr Gly Arg Arg  
65 70 75 80  
Glu Glu Thr Val Leu Ser Val Thr Val Gly Val Pro Pro Gly Gly Gln  
85 90 95  
Val Gly Gly Gly Gly Leu Ser Ala Arg Gln Ile Ile Ala Ser Cys Glu  
100 105 110  
Gly Ser Leu Arg Arg Leu Gly Val Asp His Val Asp Val Leu His Leu  
115 120 125  
Pro Arg Val Asp Arg Val Glu Pro Trp Asp Glu Val Trp Gln Ala Val  
130 135 140  
Asp Ala Leu Val Ala Ala Gly Lys Val Cys Tyr Val Gly Ser Ser Gly  
145 150 155 160  
Phe Pro Gly Trp His Ile Val Ala Ala Gln Glu His Ala Val Arg Arg  
165 170 175  
His Arg Leu Gly Leu Val Ser His Gln Cys Arg Tyr Asp Leu Thr Ser  
180 185 190  
Arg His Pro Glu Leu Glu Val Leu Pro Ala Ala Gln Ala Tyr Gly Leu  
195 200 205  
Gly Val Phe Ala Arg Pro Thr Arg Leu Gly Gly Leu Leu Gly Gly Asp  
210 215 220  
Gly Pro Gly Ala Ala Ala Ala Arg Ala Ser Gly Gln Pro Thr Ala Leu  
225 230 235 240  
Arg Ser Ala Val Glu Ala Tyr Glu Val Phe Cys Arg Asp Leu Gly Glu  
245 250 255  
His Pro Ala Glu Val Ala Leu Ala Trp Val Leu Ser Arg Pro Gly Val  
260 265 270  
Ala Gly Ala Val Val Gly Ala Arg Thr Pro Gly Arg Leu Asp Ser Ala  
275 280 285  
Leu Arg Ala Cys Gly Val Ala Leu Gly Ala Thr Glu Leu Thr Ala Leu  
290 295 300  
Asp Gly Ile Phe Pro Gly Val Ala Ala Ala Gly Ala Ala Pro Glu Ala  
305 310 315 320  
Trp Leu Arg

<210> 19  
<211> 247  
<212> PRT  
<213> Micromonospora megalomicea

<400> 19  
Met Asn Thr Trp Leu Arg Arg Phe Gly Ser Ala Asp Gly His Arg Ala  
1 5 10 15

```

Arg Leu Tyr Cys Phe Pro His Ala Gly Ala Ala Ala Asp Ser Tyr Leu
      20      25      30
Asp Leu Ala Arg Ala Leu Ala Pro Glu Val Asp Val Trp Ala Val Gln
      35      40      45
Tyr Pro Gly Arg Gln Asp Arg Arg Asp Glu Arg Ala Leu Gly Thr Ala
      50      55      60
Gly Glu Ile Ala Asp Glu Val Ala Ala Val Leu Arg Asp Leu Val Gly
      65      70      75      80
Glu Val Pro Phe Ala Leu Phe Gly His Ser Met Gly Ala Leu Val Ala
      85      90      95
Tyr Glu Thr Ala Arg Arg Leu Glu Ala Arg Pro Gly Val Arg Pro Leu
      100     105     110
Arg Leu Phe Val Ser Gly Gln Thr Ala Pro Arg Val His Glu Arg Arg
      115     120     125
Thr Asp Leu Pro Asp Glu Asp Gly Leu Val Glu Gln Met Arg Arg Leu
      130     135     140
Gly Val Ser Glu Ala Ala Leu Ala Asp Gln Gly Leu Leu Asp Met Ser
      145     150     155     160
Leu Pro Val Leu Arg Ala Asp His Arg Val Leu Arg Ser Tyr Ala Trp
      165     170     175
Gln Ala Gly Pro Pro Leu Arg Ala Gly Ile Thr Thr Leu Cys Gly Asp
      180     185     190
Thr Asp Pro Leu Thr Thr Val Glu Asp Ala Gln Arg Trp Leu Pro Tyr
      195     200     205
Ser Val Val Pro Gly Arg Thr Arg Thr Phe Pro Gly Gly His Phe Tyr
      210     215     220
Leu Ala Asp His Val Gly Glu Val Ala Glu Ser Val Ala Pro Asp Leu
      225     230     235     240
Leu Arg Leu Thr Pro Thr Gly
      245

```

&lt;210&gt; 20

&lt;211&gt; 189

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 20

```

Ile Arg Val Gln Asp Asp Asp Ala Asp Arg Leu Ser Arg Asp Glu Leu
  1      5      10      15
Thr Ser Ile Ala Leu Val Leu Leu Leu Ala Gly Phe Glu Ala Ser Val
      20      25      30
Ser Leu Ile Gly Ile Gly Thr Tyr Leu Leu Leu Thr His Pro Asp Gln
      35      40      45
Leu Ala Leu Val Arg Lys Asp Pro Ala Leu Leu Pro Gly Ala Val Glu
      50      55      60
Glu Ile Leu Arg Tyr Gln Ala Pro Pro Glu Thr Thr Thr Arg Phe Ala
      65      70      75      80
Thr Ala Glu Val Glu Ile Gly Gly Val Thr Ile Pro Ala Tyr Ser Thr
      85      90      95
Val Leu Ile Ala Asn Gly Ala Ala Asn Arg Asp Pro Gly Gln Phe Pro
      100     105     110
Asp Pro Asp Arg Phe Asp Val Thr Arg Asp Ser Arg Gly His Leu Thr
      115     120     125
Phe Gly His Gly Ile His Tyr Cys Met Gly Arg Pro Leu Ala Lys Leu
      130     135     140
Glu Gly Glu Val Ala Leu Gly Ala Leu Phe Asp Arg Phe Pro Lys Leu
      145     150     155     160
Ser Leu Gly Phe Pro Ser Asp Glu Val Val Trp Arg Arg Ser Leu Leu
      165     170     175
Leu Arg Gly Ile Asp His Leu Pro Val Arg Pro Asn Gly
      180     185

```



<210> 21  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Synthetic nucleotide DNA duplex

<400> 21  
 taagaattcg gagatctggc ctcagctcta gac 33

<210> 22  
 <211> 39  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Complementary oligo

<400> 22  
 aattgtctag agctgaggcc agatctccga attcttaat 39

<210> 23  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<400> 23  
 ttgcagcggg tgctcgggtggc ggtgcggggag gggcgctcggg tgttggggtgt ggtgggtgggt 60  
 tcggcgggtga atcaggatgg ggcgagtaat ggggttgccgg cgccgctcggg ggtggcgag 120  
 cagcgggtga ttccggcgggc gtgggggtcgt gcgggtgtgt cgggtgggga tgtgggtgtg 180  
 gtggaggcgc atgggacggg gacgcgggtt ggggatccgg tggagttggg ggcgttggtg 240  
 gggacgtatg ggggtgggtcg ggggtgggggtg ggtccgggtg tgggtgggttc ggtgaaggcg 300  
 aatgtgggtc atgtgcaggc ggcggcgggt gtgggtgggtg tgatcaagggt ggtgttgggg 360  
 ttgggtcggg ggttggtggg tccgatgggtg tgcgggggtg ggttgctcggg gttggtggat 420  
 tggtcgctcg gtgggttggg ggtggcggat ggggtgcggg ggtggccggg ggggtgtggat 480  
 ggggtgcgtc ggggtggggg gtcggcggtt ggggtgtcgg ggacgaat 528

<210> 24  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<400> 24  
 ctgcagcggg tgctcgggtggc ggtgcggggag gggcgctcggg tgttggggtgt ggtgggtgggt 60  
 tcggcgggtga atcaggatgg ggcgagtaat ggggttgccgg cgccgctcggg ggtggcgag 120  
 cagcgggtga ttccggcgggc gtgggggtcgt gcgggtgtgt cgggtgggga tgtgggtgtg 180  
 gtggaggcgc atgggacggg gacgcgggtt ggggatccgg tggagttggg ggcgttggtg 240  
 gggacgtatg ggggtgggtcg ggggtgggggtg ggtccgggtg tgggtgggttc ggtgaaggcg 300  
 aatgtgggtc atgtgcaggc ggcggcgggt gtgggtgggtg tgatcaagggt ggtgttgggg 360  
 ttgggtcggg ggttggtggg tccgatgggtg tgcgggggtg ggttgctcggg gttggtggat 420  
 tggtcgctcg gtgggttggg ggtggcggat ggggtgcggg ggtggccggg ggggtgtggat 480  
 ggggtgcgtc ggggtggggg gtcggcggtt ggggtgtcgg ggacgaat 528

<210> 25  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<220>

<221> misc\_feature

<222> (1)...(528)

<223> Sequence with codon changes as described in the  
specification at page 99, line 22 thru 101, line 23

<400> 25

ctgcagcgcc	tctccgtcgc	cgcccgcgag	ggccgcccag	tcctcggcgt	cgtcgctggc	60
tccggccgtca	accaagacgg	cgcgtaaac	ggcctcgccg	cgccctccgg	cgtcgcccag	120
cagcgcgta	tacgccgcgc	gtggggacgc	gccggagtat	cgggcggcga	cgtcggagtc	180
gtcgaaggccc	acggcaccgg	caccgcctc	ggggatcccc	tcgagctggg	cgccctcctg	240
ggcacgtacg	gcgtcggccg	cggcggcgtc	ggcccggctc	tcgtcggcag	cgtaaggcc	300
aacgtcggcc	acgtccaggc	cgcgccgggc	gtcgtcgggg	tcataaagg	cgccctcggc	360
ctcggccgcg	ggctggtcgg	cccgatggtc	tgcccgggcg	gcctcagcgg	cctcgtcgac	420
tggtcgtccg	gcggcctggg	cgtcgcggac	ggggtcgcgg	gctggccggg	cggcgtcgac	480
ggcgtccgcc	ggggcgggcg	ctcggcgctc	ggcgtcagcg	ggacgaat		528

<210> 26

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<400> 26

ggtggagtgt	gatgcggtgg	tgctcgtcgt	ggtgggggtt	tcgggtgttg	gggtgttgga	60
gggtcggtcg	ggtgcgccgt	cggttgatcg	ggtggatgtg	gtgcagccgg	tggtgttcgt	120
ggtgatggtg	tcgttggcgc	ggttggtggc	gtgggtgtgg	gttggtgcct	cggcgggtgg	180
gggtcattcg	cagggggaga	tcgcggcgcc	ggtgggtggc	gggggtgttg	cgggtgggtga	240
tggtgcgcgg	gtgggtggcg	tgccgggcgc	ggcgttgccg	gcgttgggcc	g	291

<210> 27

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<400> 27

ggtggagtgt	gatgcggtgg	tgctcgtcgt	ggtgggggtt	tcgggtgttg	gggtgttgga	60
gggtcggtcg	ggtgcgccgt	cggttgatcg	ggtggatgtg	gtgcagccgg	tggtgttcgt	120
ggtgatggtg	tcgttggcgc	ggttggtggc	gtgggtgtgg	gttggtgcct	cggcgggtgg	180
gggtcattcg	cagggggaga	tcgcggcgcc	ggtgggtggc	gggggtgttg	cgggtgggtga	240
tggtgcgcgg	gtgggtggcg	tgccgggcgc	ggcgttgccg	gcgttgggcc	g	291

<210> 28

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<220>

<221> misc\_feature

<222> (1)...(291)

<223> Sequence with codon changes as described in the  
specification at page 99, line 22 thru page 101, line 23

<400> 28

cgtggagtgc	gatgcggtcg	tgctcagcgt	cgccggcttc	agcgtgctgg	gcgtccctgga	60
gggcccgcgc	ggcgccccga	gcctggaccg	cgccgacgtg	gtccagccgg	tcctgttcgt	120
ggtcatggtc	agcctggccc	gcctgtggcg	ctgggtgcgg	gtgggtcccg	cggcgggtgg	180
cggccacagc	cagggcgaga	tcgccgcgcg	ggtcgtggcc	ggcgtccctga	gcgtcggcga	240
cggcgcccgc	gtcgtggccc	tgccgcgccg	cgccctgcgc	gccttgggcc	g	291

<210> 29

<211> 24

<212> DNA

## &lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; PCR primer

&lt;400&gt; 29

gaacaactcc tgtctgcggc cgcg

24

&lt;210&gt; 30

&lt;211&gt; 40

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; PCR primer

&lt;400&gt; 30

cggaattctc tagagtcacg tctccaaccg cttgtcgagg

40

&lt;210&gt; 31

&lt;211&gt; 51

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; PCR primer

&lt;400&gt; 31

tctagactta attaaggagg acacatatga gcgagagcag cggcatgacc g

51

&lt;210&gt; 32

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; PCR primer

&lt;400&gt; 32

aacgcctccc aggagatctc cagca

25

&lt;210&gt; 33

&lt;211&gt; 16

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligo

&lt;400&gt; 33

aattcatagc ctaggt

16

&lt;210&gt; 34

&lt;211&gt; 16

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Oligo

&lt;400&gt; 34

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number  
**WO 01/27284 A3**

(51) International Patent Classification<sup>7</sup>: C12N 15/52,  
15/53, 15/54, 15/61, 15/62, 9/04, 9/10, 9/90, C12P 19/62

(21) International Application Number: PCT/US00/27433

(22) International Filing Date: 5 October 2000 (05.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/158,305 8 October 1999 (08.10.1999) US  
60/190,024 17 March 2000 (17.03.2000) US

(71) Applicant: KOSAN BIOSCIENCES, INC. [US/US];  
3832 Bay Center Place, Hayward, CA 94545 (US).

(72) Inventors: MCDANIEL, Robert; Palo Alto, CA (US).  
VOLCHEGURKSY, Yanina; Emeryville, CA (US).

(74) Agent: FAVORITO, Carolyn, A.; Morrison & Foerster  
LLP, Suite 500, 3811 Valley Centre Drive, San Diego, CA  
92130-2332 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

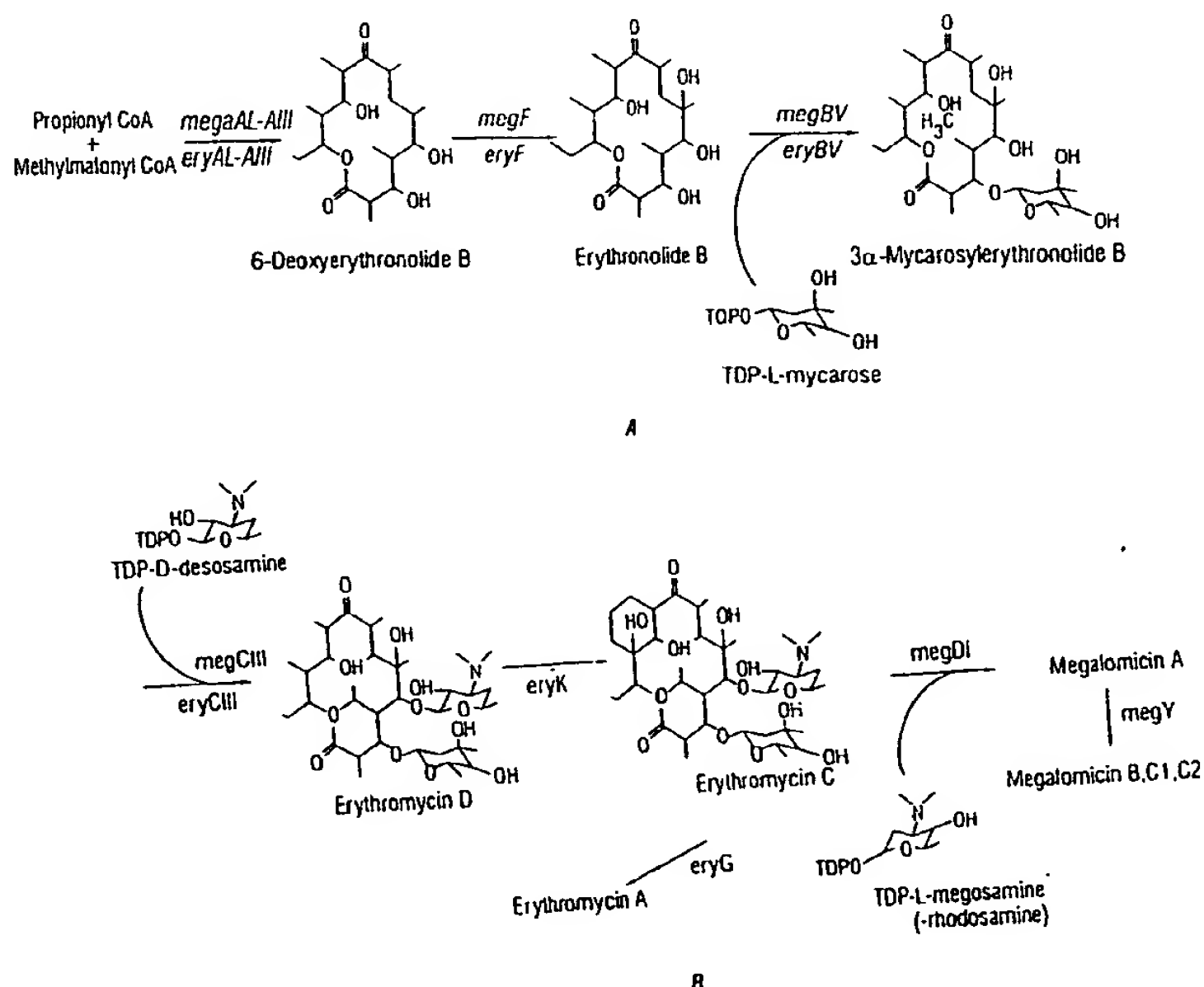
(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:  
— with international search report

(88) Date of publication of the international search report:  
28 February 2002

For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

(54) Title: RECOMBINANT MEGALOMICIN BIOSYNTHETIC GENES AND USES THEREOF



(57) Abstract: Recombinant nucleic acid, e.g. DNA compounds that encode all or a portion of the megalomicin polyketide synthase and modification enzymes are used to express recombinant polyketide synthase genes in host cells for the production of megalomicin, megalomicin derivatives, and other polyketides that are useful as antibiotics, motilides, and antiparasitics.

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/27433

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/52 C12N15/53 C12N15/54 C12N15/61 C12N15/62  
C12N9/04 C12N9/10 C12N9/90 C12P19/62

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EMBL, EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, EMBASE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 23630 A (ABBOTT LAB) 3 July 1997 (1997-07-03) the whole document claims 1-22 figures 1-3 ---	1-12, 14, 18, 19
X	WO 99 05283 A (MENDEZ CARMEN ; SALAS JOSE A (ES); RAYNAL MARIE CECILE (FR); FROMEN) 4 February 1999 (1999-02-04) the whole document claims 1-41 --- -/--	1-12, 14, 18, 19



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*G\* document member of the same patent family

Date of the actual completion of the international search

13 June 2001

Date of mailing of the international search report

09/07/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

van de Kamp, M

# INTERNATIONAL SEARCH REPORT

International Application No  
PC1/US 00/27433

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUMMERS R G ET AL.: "Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of Saccharopolyspora erythraea that are involved in L-mycarose and D-desosamine production" MICROBIOLOGY, vol. 143, 1 October 1997 (1997-10-01), pages 3251-3262, XP002061260 cited in the application abstract page 3253, right-hand column, line 47 -page 3253, left-hand column, line 19 figures 1-6; table 1 ---	1-12,14, 18,19
X	OLANO C ET AL.: "Analysis of a Streptomyces antibioticus chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring" MOLECULAR AND GENERAL GENETICS, vol. 259, no. 3, 1 August 1998 (1998-08-01), pages 299-308, XP002096258 cited in the application abstract page 300, right-hand column, line 46 -page 301, left-hand column, line 17 figures 1,2 ---	1,5-12, 19
X	XUE Y ET AL.: "A gene cluster for macrolide antibiotic biosynthesis in Streptomyces venezuelae: architecture of metabolic diversity" PROC. NATL. ACAD. SCI. USA, vol. 95, October 1998 (1998-10), pages 12111-12116, XP002166926 cited in the application abstract page 12113, left-hand column, line 4-24 figures 1,2; tables 1,2 ---	1,5-12, 19
X	OTTEN S L ET AL.: "Cloning and chracterization of the Streptomyces peucetius dmhZUV genes encoding three enzymes required for biosynthesis of the daunorubicin precursor thymidine diphospho-L-daunosamine" JOURNAL OF BACTERIOLOGY, vol. 179, no. 13, July 1997 (1997-07), pages 4446-4450, XP002166927 abstract figure 1; table 1 ---	1,5-12, 19
	-/--	



## INTERNATIONAL SEARCH REPORT

International Application No

PCI/US 00/27433

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>OTTEN S L ET AL.: "Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunosamine biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis" JOURNAL OF BACTERIOLOGY, vol. 177, no. 22, November 1995 (1995-11), pages 6688-6692, XP002166928 abstract figure 1</p> <p>---</p>	1,5-12, 19
X	<p>TORKKELL S ET AL.: "Characterization of Streptomyces nogalater genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis" MOLECULAR AND GENERAL GENETICS, vol. 256, no. 2, September 1997 (1997-09), pages 203-209, XP002166929 cited in the application abstract figure 1</p> <p>---</p>	1,5-12, 19
A	<p>SWAN D G ET AL.: "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence" MOLECULAR AND GENERAL GENETICS, vol. 242, no. 3, 1994, pages 358-362, XP002087278 cited in the application abstract page 358, right-hand column, line 5 -page 361, left-hand column, line 18</p> <p>---</p>	1,9
Y	<p>US 3 819 611 A (WEINSTEIN M ET AL) 25 June 1974 (1974-06-25) the whole document</p> <p>---</p>	1-12,14, 18-20
Y	<p>MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE, vol. 325, 26 February 1987 (1987-02-26), pages 818-821, XP002075972 abstract</p> <p>---</p>	1-12,14, 18-20
A	<p>NAKAGAWA A ET AL.: "Structure and stereochemistry of macrolides" MACROLIDE ANTIBIOTICS. OMURA S (ED.). PUBLISHER: ACADEMIC, ORLANDO, FLORIDA, 1984, pages 37-84, XP001006199 page 46, line 25 -page 48, line 4</p> <p>---</p> <p>-/--</p>	

# INTERNATIONAL SEARCH REPORT

International Application No  
PC1/US 00/27433

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEONARD KATZ: "Manipulation of modular polyketide synthases" CHEMICAL REVIEWS, vol. 97, no. 7, 1997, pages 2557-2575, XP002103748 the whole document	1-12,14, 18,19
A	LIU H -W ET AL: "Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria" ANNUAL REVIEW OF MICROBIOLOGY, vol. 48, 1994, pages 223-256, XP002061259 page 234, line 24 -page 237, line 9; figures 8,9	1,5-12, 19
A	CARRERAS C W ET AL.: "Engineering of modular polyketide synthases to produce novel polyketides" CURRENT OPINION IN BIOTECHNOLOGY, vol. 9, no. 4, August 1998 (1998-08), pages 403-411, XP000993508 the whole document	14,18
A	HUTCHINSON C R: "Combinatorial biosynthesis for new drug discovery" CURRENT OPINION IN MICROBIOLOGY, vol. 1, no. 3, June 1998 (1998-06), pages 319-329, XP000993550 the whole document	14,18
A	MCDANIEL R ET AL.: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 96, 1999, pages 1846-1851, XP000910246 cited in the application abstract	14,18
P,X	VOLCHEGURSKY Y ET AL.: "Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in Saccharopolyspora erythrea" MOLECULAR MICROBIOLOGY, vol. 37, no. 4, August 2000 (2000-08), pages 752-762, XP002166930 the whole document	1-6, 8-13, 18-20
P,X	WO 00 00500 A (LEADLAY PETER FRANCIS ;CORTES JESUS (GB); STAUNTON JAMES (GB); BIO) 6 January 2000 (2000-01-06) claim 24	14
	---	
	--- -/--	

INTERNATIONAL SEARCH REPORT

International Application No  
PC1/US 00/27433

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 00 63361 A (KOSAN BIOSCIENCES INC) 26 October 2000 (2000-10-26) page 9, line 3-9 page 14, line 26 -page 16, line 2 claim 3 -----	1-13, 18-20

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCI/US 00/27433

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9723630 A	03-07-1997	US 5998194 A EP 0874548 A JP 2000502899 T	07-12-1999 04-11-1998 14-03-2000
WO 9905283 A	04-02-1999	FR 2766496 A FR 2786200 A EP 1032679 A	29-01-1999 26-05-2000 06-09-2000
US 3819611 A	25-06-1974	BE 715638 A CA 931891 A CH 534206 A CS 157635 B DE 1767565 A DK 123422 B ES 354296 A FI 46519 B FR 8066 M GB 1229835 A IE 31918 B IL 30067 A LU 56131 A NL 6807363 A NO 128225 B OA 4027 A SE 349323 B	25-11-1968 14-08-1973 28-02-1973 16-09-1974 14-10-1971 19-06-1972 16-10-1969 02-01-1973 06-07-1970 28-04-1971 07-02-1973 28-09-1972 11-09-1968 27-11-1968 15-10-1973 15-09-1979 25-09-1972
WO 0000500 A	06-01-2000	AU 4524599 A AU 4524799 A BR 9911710 A BR 9911712 A EP 1091971 A EP 1090123 A WO 0000618 A	17-01-2000 17-01-2000 20-03-2001 20-03-2001 18-04-2001 11-04-2001 06-01-2000
WO 0063361 A	26-10-2000	AU 4241800 A	02-11-2000

CORRECTED VERSION

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
19 April 2001 (19.04.2001)

PCT

(10) International Publication Number  
WO 01/27284 A3

(51) International Patent Classification<sup>7</sup>: C12N 15/52,  
15/53, 15/54, 15/61, 15/62, 9/04, 9/10, 9/90, C12P 19/62

(72) Inventors: MCDANIEL, Robert; Palo Alto, CA (US).  
VOLCHEGURKSY, Yanina; Emeryville, CA (US).

(21) International Application Number: PCT/US00/27433

(74) Agent: FAVORITO, Carolyn, A.; Morrison & Foerster  
LLP, Suite 500, 3811 Valley Centre Drive, San Diego, CA  
92130-2332 (US).

(22) International Filing Date: 5 October 2000 (05.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/158,305 8 October 1999 (08.10.1999) US  
60/190,024 17 March 2000 (17.03.2000) US

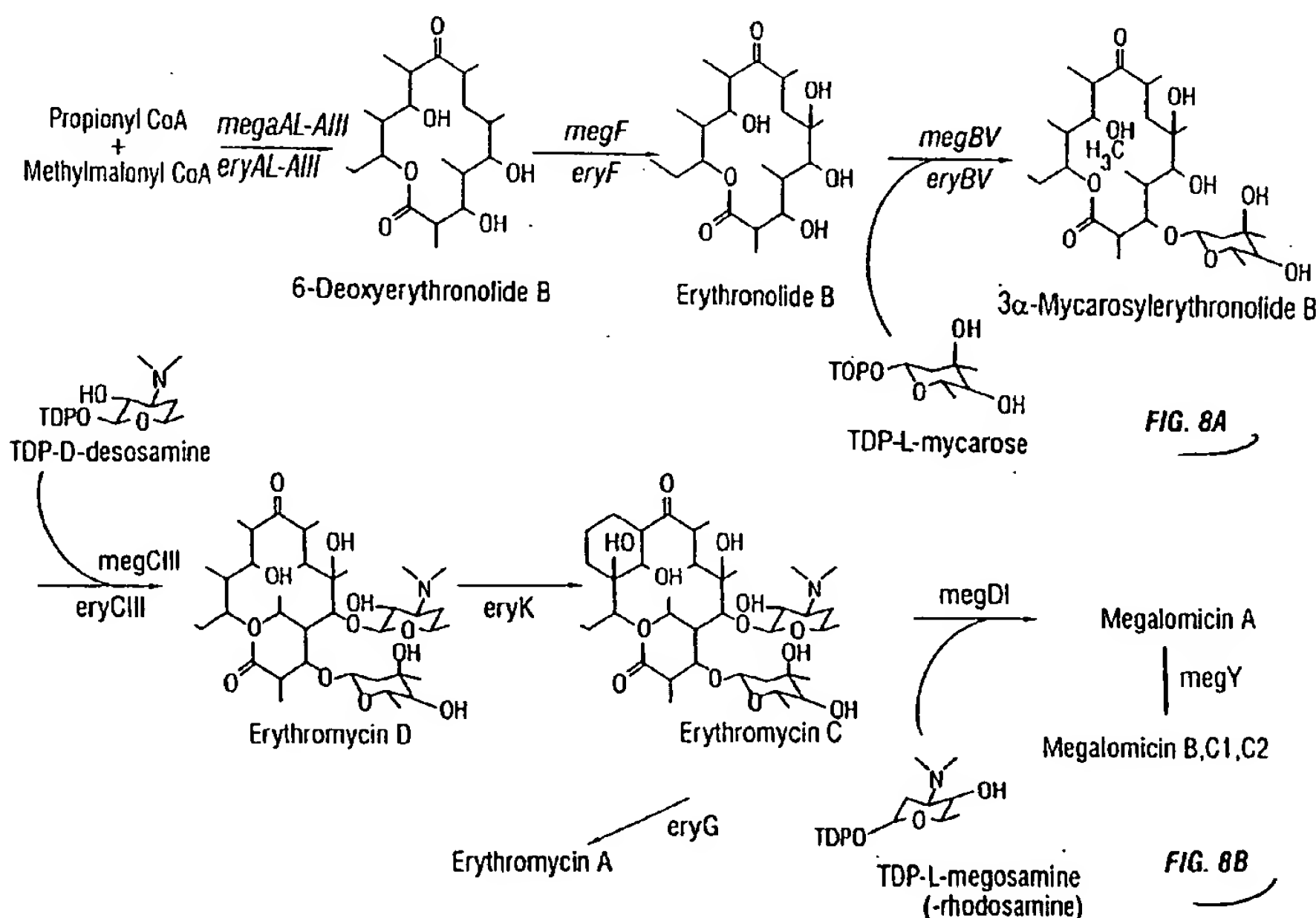
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(71) Applicant: KOSAN BIOSCIENCES, INC. [US/US];  
3832 Bay Center Place, Hayward, CA 94545 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW); Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM); European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: RECOMBINANT MEGALOMICIN BIOSYNTHETIC GENES AND USES THEREOF



(57) Abstract: Recombinant nucleic acid, e.g. DNA compounds that encode all or a portion of the megalomicin polyketide synthase and modification enzymes are used to express recombinant polyketide synthase genes in host cells for the production of megalomicin, megalomicin derivatives, and other polyketides that are useful as antibiotics, motilides, and antiparasitics.



IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(48) Date of publication of this corrected version:

28 March 2002

**Published:**

— *with international search report*

(15) Information about Correction:

see PCT Gazette No. 13/2002 of 28 March 2002, Section II

(88) Date of publication of the international search report:  
28 February 2002

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*



### Title

## **Recombinant Megalomicin Biosynthetic Genes And Uses Thereof**

### Cross-Reference to Priority Application

5        This application claims priority to provisional U.S. patent application  
Serial No. 60/158,305, filed 8 October 1999, and provisional U.S. patent  
application Serial No. 60/190,024, filed 17 March 2000 under 35 U.S.C. § 119(e).  
The content of the above referenced applications is incorporated herein by  
reference in its entirety.

10

### Field of the Invention

The present invention provides recombinant methods and materials for  
producing polyketides by recombinant DNA technology. The invention relates to  
15 the fields of agriculture, animal husbandry, chemistry, medicinal chemistry,  
medicine, molecular biology, pharmacology, and veterinary technology.

### Background of the Invention

Polyketides represent a large family of diverse compounds synthesized  
20 from 2-carbon units through a series of condensations and subsequent  
modifications. Polyketides occur in many types of organisms, including fungi and  
mycelial bacteria, in particular, the actinomycetes. There are a wide variety of  
polyketide structures, and the class of polyketides encompasses numerous  
compounds with diverse activities. Erythromycin, FK-506, FK-520, megalomicin,  
25 narbomycin, oleandomycin, picromycin, rapamycin, spinocyn, and tylosin are  
examples of such compounds. Given the difficulty in producing polyketide  
compounds by traditional chemical methodology, and the typically low production  
of polyketides in wild-type cells, there has been considerable interest in finding  
improved or alternate means to produce polyketide compounds. See PCT  
30 publication Nos. WO 93/13663; WO 95/08548; WO 96/40968; WO 97/02358;  
and WO 98/27203; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837;  
5,149,639; 5,672,491; and 5,712,146; Fu *et al.*, 1994, *Biochemistry* 33: 9321-  
9326; McDaniel *et al.*, 1993, *Science* 262: 1546-1550; and Rohr, 1995, *Angew.*

*Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by polyketide synthase (PKS) enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKS enzymes are encoded by PKS genes that usually consist of three or more open reading frames (ORFs). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

Modular PKSs are responsible for producing a large number of 12-, 14-, and 16-membered macrolide antibiotics including erythromycin, megalomicin, methymycin, narbomycin, oleandomycin, picromycin, and tylosin. Each ORF of a modular PKS can comprise one, two, or more "modules" of ketosynthase activity, each module of which consists of at least two (if a loading module) and more typically three (for the simplest extender module) or more enzymatic activities or "domains." These large multifunctional enzymes (>300,000 kDa) catalyze the biosynthesis of polyketide macrolactones through multistep pathways involving decarboxylative condensations between acyl thioesters followed by cycles of varying  $\beta$ -carbon processing activities (see O'Hagan, D. *The polyketide metabolites*; E. Horwood: New York, 1991, incorporated herein by reference).

During the past half decade, the study of modular PKS function and specificity has been greatly facilitated by the plasmid-based *Streptomyces coelicolor* expression system developed with the 6-deoxyerythronolide B (6-dEB) synthase (DEBS) genes (see Kao *et al.*, 1994, *Science*, 265: 509-512, McDaniel *et al.*, 1993, *Science* 262: 1546-1557, and U.S. Patent Nos. 5,672,491 and 5,712,146, each of which is incorporated herein by reference). The advantages to this plasmid-based genetic system for DEBS are that it overcomes the tedious and limited techniques for manipulating the natural DEBS host organism, *Saccharopolyspora erythraea*, allows more facile construction of recombinant PKSs, and reduces the complexity of PKS analysis by providing a "clean" host background. This system also expedited construction of the first combinatorial

modular polyketide library in *Streptomyces* (see PCT publication No. WO 98/49315, incorporated herein by reference).

5 The ability to control aspects of polyketide biosynthesis, such as monomer selection and degree of  $\beta$ -carbon processing, by genetic manipulation of PKSs has stimulated great interest in the combinatorial engineering of novel antibiotics (see Hutchinson, 1998, *Curr. Opin. Microbiol.* 1: 319-329; Carreras and Santi, 1998, *Curr. Opin. Biotech.* 9: 403-411; and U.S. Patent Nos. 5,712,146 and 5,672,491, each of which is incorporated herein by reference). This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that  
10 encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known  
15 PKS gene clusters.

Megalomicin is a macrolide antibiotic produced by *Micromonospora megalomicea*, a member of the Actinomycetales family of soil bacteria that produces many types of biologically active compounds. Megalomicin is a glycoside of erythromycin A, a widely used antibacterial drug with little or no  
20 antimalarial activity. Megalomicin has antibacterial properties similar to those of erythromycin, and in 1998, it was discovered also to have potent antiparasitic activity and low toxicity. The antiparasitic activity may be related to the effect megalomicin has on protein trafficking in eukaryotes, where it appears to inhibit vesicular transport between the medial and trans-Golgi, resulting in under-  
25 sialylation of proteins. Hence, megalomicin offers an exciting opportunity to develop a new class of antiparasitic drugs with a different mechanism of action than the drugs currently in use and, therefore, possibly active against drug-resistant forms of *Plasmodium falciparum*.

The number and diversity of megalomicin derivatives have been limited  
30 due to the inability to manipulate the PKS genes, which have not previously been available in recombinant form. Genetic systems that allow rapid engineering of the megalomicin biosynthetic genes would be valuable for creating novel compounds for pharmaceutical, agricultural, and veterinary applications. The production of

such compounds could be more readily accomplished if the heterologous expression of the megalomicin biosynthetic genes in *Streptomyces coelicolor* and *S. lividans* and other host cells were possible. The present invention meets these and other needs.

5

#### Summary of the Invention

The present invention provides recombinant methods and materials for expressing PKS enzymes and polyketide modification enzymes derived in whole and in part from the megalomicin biosynthetic genes in recombinant host cells.

- 10 The invention also provides the polyketides produced by such PKS enzymes. The invention provides in recombinant form all of the genes for the proteins that constitute the complete PKS that ultimately results, in *Micromonospora megalomicea*, in the production of megalomicin. Thus, in one embodiment, the invention is directed to recombinant materials comprising nucleic acids with
- 15 nucleotide sequences encoding at least one domain, module, or protein encoded by a megalomicin PKS gene. In one preferred embodiment of the invention, the DNA compounds of the invention comprise a coding sequence for at least one and preferably two or more of the domains of the loading module and extender modules 1 through 6, inclusive, of the megalomicin PKS.

- 20 In one embodiment, the invention provides a recombinant expression vector that comprises a heterologous promoter positioned to drive expression of one or more of the megalomicin biosynthetic genes. In a preferred embodiment, the promoter is derived from another PKS gene. In a related embodiment, the invention provides recombinant host cells comprising one or more expression
- 25 vectors that produce(s) megalomicin or a megalomicin derivative or precursor. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In another embodiment, the invention provides a recombinant expression vector that comprises a promoter positioned to drive expression of a hybrid PKS comprising all or part of the megalomicin PKS and at least a part of a second PKS.

- 30 In a related embodiment, the invention provides recombinant host cells comprising the vector that produces the hybrid PKS and its corresponding polyketide. In a preferred embodiment, the host cell is *Streptomyces lividans* or *S. coelicolor*.

In a related embodiment, the invention provides recombinant materials for the production of libraries of polyketides wherein the polyketide members of the library are synthesized by hybrid PKS enzymes of the invention. The resulting polyketides can be further modified to convert them to other useful compounds, such as antibiotics, motilides, and antiparasitics, typically through hydroxylation and/or glycosylation. Modified macrolides provided by the invention that are useful intermediates in the preparation of antiparasitics are of particular benefit.

In another related embodiment, the invention provides a method to prepare a nucleic acid that encodes a modified PKS, which method comprises using the megalomicin PKS encoding sequence as a scaffold and modifying the portions of the nucleotide sequence that encode enzymatic activities, either by mutagenesis, inactivation, deletion, insertion, or replacement. The thus modified megalomicin PKS encoding nucleotide sequence can then be expressed in a suitable host cell and the cell employed to produce a polyketide different from that produced by the megalomicin PKS. In addition, portions of the megalomicin PKS coding sequence can be inserted into other PKS coding sequences to modify the products thereof.

In another related embodiment, the invention is directed to a multiplicity of cell colonies, constituting a library of colonies, wherein each colony of the library contains an expression vector for the production of a modular PKS derived in whole or in part from the megalomicin PKS. Thus, at least a portion of the modular PKS is identical to that found in the PKS that produces megalomicin and is identifiable as such. The derived portion can be prepared synthetically or directly from DNA derived from organisms that produce megalomicin. In addition, the invention provides methods to screen the resulting polyketide and antibiotic libraries.

The invention also provides novel polyketides, motilides, antibiotics, antiparasitics and other useful compounds derived therefrom. The compounds of the invention can also be used in the manufacture of another compound. In a preferred embodiment, the compounds of the invention are formulated in a mixture or solution for administration to an animal or human.

In a specific embodiment, the invention provides an isolated nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The isolated

nucleic acid fragment can be a DNA or a RNA. Preferably, the isolated nucleic acid fragment is a recombinant DNA compound.

The isolated nucleic acid fragment can comprise a single, multiple or all the open reading frame(s) (ORF) of the megalomicin PKS or a megalomicin modification enzyme. Exemplary ORFs of megalomicin PKS include the ORFs of the *megAI*, *megAII* and *megAIII* genes. The isolated nucleic acid fragment can also encode a single, multiple, or all of the domains of the megalomicin PKS. Exemplary domains of the megalomicin PKS include a TE domain, a KS domain, an AT domain, an ACP domain, a KR domain, a DH domain and an ER domain.

10 In a preferred embodiment, the nucleic acid fragment encodes a module of the megalomicin PKS. In another preferred embodiment, the nucleic acid fragment encodes the loading module, a thioesterase domain, and all six extender modules of the megalomicin PKS.

Megalomicin modification enzymes include those enzymes involved in the conversion of 6-dEB into a megalomicin such as the enzymes encoded by the *megF*, *megBV*, *megCIII*, *megK*, *megDI* and *megG* (renamed *megY*) genes.

15 Megalomicin modification enzymes also include those enzymes involved in the biosynthesis of mycarose, megosamine or desosamine, which are used as biosynthetic intermediates in the biosynthesis of various megalomicin species and other related polyketides. The enzymes that are involved in biosynthesis of mycarose, megosamine or desosamine are described in Figures 5 and 10.

In a preferred embodiment, the invention provides an isolated nucleic acid fragment which hybridizes to a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1, under low, medium or high stringency. More preferably, the nucleic acid fragment comprises, consists or consists essentially of a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1.

25

In another specific embodiment, the invention provides a substantially purified polypeptide, which is encoded by a nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The polypeptide can comprise a single domain, multiple domains or a full-length megalomicin PKS or megalomicin modification enzyme. Functional fragments, analogs or derivatives of the megalomicin PKS or megalomicin modification enzyme polypeptides are

30



also provided. Preferably, such fragments, analogs or derivatives can be recognized by an antibody raised against a megalomicin PKS or megalomicin modification enzyme. Also preferably, such fragments, analogs or derivatives comprise an amino acid sequence that has at least 60% identity, more preferably at least 90% identity, to their wild type counterparts.

In still another specific embodiment, the invention provides an antibody, or a fragment or derivative thereof, which immuno-specifically binds to a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The antibody can be a monoclonal or polyclonal antibody or an antibody fragment.

10 Preferably, the antibody is a monoclonal antibody.

In yet another specific embodiment, the invention provides a recombinant DNA expression vector comprising the recombinant DNA compound encoding at least a domain of the megalomicin PKS or a megalomicin modification enzyme, wherein said domain is operably linked to a promoter. Preferably, the recombinant DNA expression vector further comprises an origin of replication or a segment of DNA that enables chromosomal integration.

In yet another specific embodiment, the invention provides a recombinant host cell comprising the above-described recombinant DNA expression vector encoding at least a domain of megalomicin PKS or the megalomicin modification enzyme. The recombinant host cells can be any suitable host cells including animal, mammalian, plant, fungal, yeast, and bacterial cells. Preferably, the recombinant host cells are *Streptomyces* cells, such as *Streptomyces lividans* and *S. coelicolor* cells, or *ccharopolyspora* cells, such as *Saccharopolyspora erythraea* cells. Also preferably, the recombinant host cells do not produce megalomicin in their untransformed, non-recombinant state.

When the recombinant host cell contains nucleic acid encoding more than one megalomicin PKS or megalomicin modification enzyme, or domains thereof, such nucleic acid material can be located at a single genetic locus, *e.g.*, on a single plasmid or at a single chromosomal locus, or at different genetic loci, *e.g.*, on separate plasmids and/or chromosomal loci. In one example, the invention provides a recombinant host cell, which comprises at least two separate autonomously replicating recombinant DNA expression vectors, and each of said vectors comprises a recombinant DNA compound encoding a megalomicin PKS

domain or a megalomicin modification enzyme operably linked to a promoter. In another example, the invention provides a recombinant host cell, which comprises at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant DNA compound encoding a megalomicin  
5 PKS domain or a megalomicin modification enzyme operably linked to a promoter. Preferably, the autonomously replicating recombinant DNA expression vector and/or the modified chromosome further comprises distinct selectable markers.

10 In a preferred embodiment, the cell comprises three different vectors, one of which is integrated into the chromosome and two of which are autonomously replicating, and each of the vectors comprises a *meg* PKS gene. Optionally, one or more of the *meg* PKS genes contains one or more domain alterations, such as a deletion or substitution of a *meg* PKS domain with a domain from another PKS.

15 In yet another specific embodiment, the invention provides a hybrid PKS, which is produced from a recombinant gene that comprises at least a portion of a megalomicin PKS gene and at least a portion of a second PKS gene for a polyketide other than megalomicin. For example, and without limitation, the second PKS gene can be a narbonolide PKS gene, an oleandolide PKS gene, or a  
20 rapamycin PKS gene. In one embodiment, the hybrid PKS is composed of a loading module and six extender modules, wherein at least one domain of any one of extender modules 1 through 6, inclusive, is a domain of an extender module of megalomicin PKS. In another preferred embodiment, the hybrid PKS comprises a megalomicin PKS that has a non-functional KS domain in module 1.

25 In yet another specific embodiment, the invention provides a method of producing a polyketide, which method comprises growing the recombinant host cell comprising a recombinant DNA expression vector encoding at least a domain of the megalomicin PKS or a megalomicin modification enzyme under conditions whereby the megalomicin PKS domain or the megalomicin modification enzyme  
30 comprised by the recombinant expression vector is produced and the polyketide is synthesized by the cell, and recovering the synthesized polyketide. Preferably, the recombinant host cell comprises a recombinant expression vector that encodes at least a portion of a *megAI*, *megAII*, or *megAIII* gene.

These and other embodiments of the invention are described in more detail in the following description, the examples, and claims set forth below.

### Brief Description of the Figures

5           Figure 1 shows restriction site and function maps of the insert DNA in cosmids pKOS079-138B, pKOS079-93D, pKOS079-93A, and pKOS079-124B of the invention. Various restriction sites (*Xho*I, *Bgl*II, *Nsi*I) are also shown. The location of the megalomicin biosynthetic genes is shown below the solid lines indicating the cosmid inserts. The genes are shown as arrows pointing in the  
10       direction of transcription. The approximate size (in kilobase (kb) pairs) of the gene cluster is indicated in 5000 bp (i.e., 5K, 10K, and the like.) increments on a solid bar beneath the arrows indicating the genes.

          Figure 2 shows a more detailed map of the megalomicin biosynthetic gene cluster. The various open reading frames are shown as arrows pointing in the  
15       direction of transcription. A line indicates the size in base pairs (in 1000 bp increments) of the gene cluster. The various domains of the megalomicin PKS are also shown. Other genes of the megalomicin biosynthetic gene cluster not shown in this Figure are located in the insert DNA of cosmids pKOS0138B and pKOS0124B.

20           Figure 3 shows the structures of the megalomicins, azithromycin and erythromycin A.

          Figure 4 shows the modules and domains of DEBS and the megalomicin PKS.

          Figure 5 shows the compounds and reactions in the erythromycin  
25       biosynthetic pathway and also for megalomicin biosynthesis. Genes that produce the various enzymes that catalyze each of the steps in the biosynthetic pathway are indicated.

          Figure 6 shows the biosynthetic pathway for the formation of desosamine, rhodosamine, and mycarose, as well as the genes that produce the various enzymes  
30       that catalyze each of the steps in the biosynthetic pathway.

          Figure 7 depicts nucleotide and amino acid sequence of *Micromonospora megalomicea* megalomicin biosynthetic genes (GenBank Accession No. AF263245, incorporated herein by reference).

Figure 8 depicts the biosynthesis of the erythromycins and megalomicins and the enzymes that mediate the biosynthesis of each.

Figure 9 depicts the cloned megalomicin biosynthetic gene cluster and certain cosmids of the invention that comprise portions of the cluster.

5        Figure 10 depicts the biosynthesis of megosamine, mycarose, and desosamine.

#### Detailed Description of the Invention

The present invention provides useful compounds and methods for  
10    producing polyketides in recombinant host cells. As used herein, the term recombinant refers to a compound or composition produced by human intervention. The invention provides recombinant DNA compounds encoding all or a portion of the megalomicin biosynthetic genes. The invention provides recombinant expression vectors useful in producing the megalomicin PKS and  
15    hybrid PKSs composed of a portion of the megalomicin PKS in recombinant host cells. The invention also provides the polyketides produced by the recombinant PKS and polyketide modification enzymes.

To appreciate the many and diverse benefits and applications of the invention, the description of the invention below is organized as follows. In  
20    Section I, common definitions used throughout this application are provided. In Section II, structural and functional characteristics of megalomicin are described. In Section III, the recombinant megalomicin biosynthetic genes and other recombinant nucleic acids provided by the invention are described. In Section IV, polypeptides and proteins encoded by the megalomicin biosynthetic genes and  
25    antibodies that specifically bind to such polypeptides and proteins provided by the invention are described. In Section V, methods for heterologous expression of the megalomicin biosynthetic genes provided by the invention are described. In Section VI, the hybrid PKS genes provided by the invention are described. In Section VII, host cells containing multiple megalomicin biosynthetic genes and  
30    nucleic acid fragments on separate express vectors provided by the invention are described. In Section VIII, the polyketide compounds provided by the invention and pharmaceutical compositions of those compounds are described. The detailed description is followed by working examples illustrating the invention.

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which this invention belongs. All patents, applications, published applications and other publications and sequences from GenBank and other data  
5 bases referred to herein are incorporated by reference in their entirety.

### Section I. Definitions

As used herein, domain refers to a portion of a molecule, *e.g.*, proteins or nucleic acids, that is structurally and/or functionally distinct from another portion  
10 of the molecule.

As used herein, antibody includes antibody fragments, such as Fab fragments, which are composed of a light chain and the variable region of a heavy chain.

As used herein, biological activity refers to the *in vivo* activities of a  
15 compound or physiological responses that result upon *in vivo* administration of a compound, composition or other mixture. Biological activity, thus, encompasses therapeutic effects and pharmaceutical activity of such compounds, compositions and mixtures. Biological activities may be observed in in vitro systems designed to test or use such activities.

20 As used herein, a combination refers to any association between two or among more items.

As used herein, a composition refers to any mixture. It may be a solution, a suspension, liquid, powder, a paste, aqueous, non-aqueous or any combination thereof.

25 As used herein, derivative or analog of a molecule refers to a portion derived from or a modified version of the molecule.

As used herein, operably linked, operatively linked or operationally associated refers to the functional relationship of DNA with regulatory and effector sequences of nucleotides, such as promoters, enhancers, transcriptional  
30 and translational stop sites, and other signal sequences. For example, operative linkage of DNA to a promoter refers to the physical and functional relationship between the DNA and the promoter such that the transcription of such DNA is initiated from the promoter by an RNA polymerase that specifically recognizes,

binds to and transcribes the DNA. To optimize expression and/or *in vitro* transcription, it may be helpful to remove, add or alter 5' untranslated portions of the clones to eliminate extra, potentially inappropriate alternative translation initiation (*i.e.*, start) codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites (see, *e.g.*, Kozak, *J. Biol. Chem.*, 266:19867-19870 (1991)) can be inserted immediately 5' of the start codon and may enhance expression. The desirability of (or need for) such modification may be empirically determined.

10       As used herein, pharmaceutically acceptable salts, esters or other derivatives of the conjugates include any salts, esters or derivatives that may be readily prepared by those of skill in this art using known methods for such derivatization and that produce compounds that may be administered to animals or humans without substantial toxic effects and that either are pharmaceutically  
15       active or are prodrugs.

          As used herein, a promoter region or promoter element refers to a segment of DNA or RNA that controls transcription of the DNA or RNA to which it is operatively linked. The promoter region includes specific sequences that are sufficient for RNA polymerase recognition, binding and transcription initiation.  
20       This portion of the promoter region is referred to as the promoter. In addition, the promoter region includes sequences that modulate this recognition, binding and transcription initiation activity of RNA polymerase. These sequences may be *cis* acting or may be responsive to *trans* acting factors. Promoters, depending upon the nature of the regulation, may be constitutive or regulated.

25       As used herein: stringency of hybridization in determining percentage mismatch is as follows: (1) high stringency: 0.1 x SSPE, 0.1% SDS, 65°C; (2) medium stringency: 0.2 x SSPE, 0.1% SDS, 50°C; and (3) low stringency: 1.0 x SSPE, 0.1% SDS, 50°C. Equivalent stringencies may be achieved using alternative buffers, salts and temperatures.



The term substantially identical or homologous or similar varies with the context as understood by those skilled in the relevant art and generally means at least 70%, preferably means at least 80%, more preferably at least 90%, and most preferably at least 95% identity.

5           As used herein, substantially identical to a product means sufficiently similar so that the property of interest is sufficiently unchanged so that the substantially identical product can be used in place of the product.

          As used herein, isolated means that a substance is either present in a preparation at a concentration higher than that substance is found in nature or in its  
10 naturally occurring state or that the substance is present in a preparation that contains other materials with which the substance is not associated with in nature. As an example of the latter, an isolated meg PKS protein includes a meg PKS protein expressed in a *Streptomyces coelicolor* or *S. lividans* host cell.

          As used herein, substantially pure means sufficiently homogeneous to  
15 appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as enzymatic and  
20 biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to those of skill in the art. A substantially chemically pure compound may, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, vector or plasmid refers to discrete elements that are used to introduce heterologous DNA into cells for either expression or replication thereof. Selection and use of such vehicles are well known within the skill of the artisan. An expression vector includes vectors capable of expressing DNAs that are operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

## Section II. Megalomicins

The megalomicins were discovered in 1969 at Schering Corp. as antibacterial agents produced by *Micromonospora megalomicea* (see Weinstein *et al.*, 1969, *J. Antibiotics* 22: 253-258, and U.S. Patent No. 3,632,750, both of which are incorporated herein by reference). Although the initial structural assignment was in error, a thorough reassessment of NMR data coupled with an X-ray crystal structure of a megalomicin A derivative (see Nakagawa and Omura, "Structure and Stereochemistry of Macrolides" in *Macrolide Antibiotics* (S. Omura, ed.), Academic Press, NY, 1984, incorporated herein by reference) established the structures shown in Figure 3. The megalomicins are 6-O-glycosides of erythromycin C with acetyl or propionyl groups esterified at the 3''' or 4''' hydroxyls of the mycarose sugar at the C-3-position. The C-6 sugar has been named "megosamine," although it had been identified 5 to 10 years earlier as L-rhodamine or N-dimethyldaunosamine, deoxyamino sugars commonly present in the anthracycline antitumor drugs. The antibacterial potency, spectrum of activity, and toxicity (LD<sub>50</sub> acute, 7-7.5 g/kg s.c. or oral; subacute, >500 mg/kg) of the megalomicins is similar to that of erythromycin A.

The megalomicins have two modes of biological activity. As antibacterials, they act like the erythromycins, which inhibit protein synthesis at the translocation step by selective binding to the bacterial 50S ribosomal RNA. They also affect

protein trafficking in eukaryotic cells (see Bonay *et al.*, 1996, *J. Biol. Chem.* 271:3719-3726, incorporated herein by reference). Although the mechanism of action is not entirely clear, it appears to involve inhibition of vesicular transport between the medial and trans Golgi, resulting in under-sialylation of proteins. The megalomicins also strongly inhibit the ATP-dependent acidification of lysosomes *in vivo* (see Bonay *et al.*, 1997, *J. Cell. Sci.* 110:1839-1849, incorporated herein by reference) and cause an anomalous glycosylation of viral proteins, which may be responsible for their antiviral activity against herpes (Tox<sub>50</sub>, 70-100  $\mu$ M; see Alarcon *et al.*, 1984, *Antivir. Res.* 4:231-243, and Alarcon *et al.*, 1988, *FEBS Lett.* 231:207-211, both of which are incorporated herein by reference).

Strikingly, the megalomicins are potent antiparasitic agents, showing an IC<sub>50</sub> of 1  $\mu$ g/ml in blocking intracellular replication of *Plasmodium falciparum* infected erythrocytes (see Bonay *et al.*, 1998, *Antimicrob. Agents Chemother.* 42:2668-2673, incorporated herein by reference). The megalomicins are effective against *Trypanosoma cruzi* and *T. brucei* (IC<sub>50</sub>, 0.2-2  $\mu$ g/ml) plus *Leishmania donovani* and *L. major* promastigotes (IC<sub>50</sub>, 3 and 8  $\mu$ g/ml, respectively). Megalomicin is also active against the intracellular replicative, amastigote form of *T. cruzi*, completely preventing its replication in infected murine LLC/MK2 macrophages at a dose of 5  $\mu$ g/ml. Importantly, the effective drug concentration is 500-fold less than the acute LD<sub>50</sub> in mammals, and there is no toxicity to BALB/c mice at doses (50 mg/kg) that are completely curative for *T. brucei* infections. Because the erythromycins do not have such activity, although azithromycin (Figure 3) has been reported to be an effective acute and prophylactic treatment for malaria caused by *P. vivax* and *P. falciparum* (see Taylor *et al.*, 1999, *Clin. Infect. Dis.* 28:74-81, incorporated herein by reference), the antiparasitic action of the megalomicins is unique and probably related to the presence of the deoxyamino sugar megosamine at C-6 (Figure 3). Consequently, the megalomicins could be developed into potent antimalarial drugs with a high therapeutic index and be active against *P. falciparum* and other species that are resistant to currently used classes of antimalarials. They also could lead to potent antiparasitic agents against leishmaniasis, trypanosomiasis, and Chagas' disease. In view of the widespread use of the erythromycins and their good oral availability plus the low mammalian toxicity of macrolides in general, the megalomicins could be used prophylactically

to combat malaria, and as fermentation products, the megalomicins should be relatively inexpensive to produce.

The megalomicins belong to the polyketide class of natural products whose members have diverse structural and pharmacological properties (see Monaghan and Tkacz, 1990, *Annu. Rev. Microbiol.* 44: 271, incorporated herein by reference). The megalomicins are assembled by polyketide synthases through successive condensations of activated coenzyme-A thioester monomers derived from small organic acids such as acetate, propionate, and butyrate. Active sites required for condensation include an acyltransferase (AT), acyl carrier protein (ACP), and beta-ketoacylsynthase (KS). Each condensation cycle results in a  $\beta$ -keto group that undergoes all, some, or none of a series of processing activities. Active sites that perform these reactions include a ketoreductase (KR), dehydratase (DH), and enoylreductase (ER). Thus, the absence of any beta-keto processing domain results in the presence of a ketone, a KR alone gives rise to a hydroxyl, a KR and DH result in an alkene, while a KR, DH, and ER combination leads to complete reduction to an alkane. After assembly of the polyketide chain, the molecule typically undergoes cyclization(s) and post-PKS modification (e.g. glycosylation, oxidation, acylation) to achieve the final active compound.

Macrolides such as erythromycin and megalomicin are synthesized by modular PKSs (see Cane *et al.*, 1998, *Science* 282: 63, incorporated herein by reference). For illustrative purposes, the PKS that produces the erythromycin polyketide (6-deoxyerythronolide B synthase or DEBS; see U.S. Patent No. 5,824,513, incorporated herein by reference) is shown in Figure 4. DEBS is the most characterized and extensively used modular PKS system. DEBS is particularly relevant to the present invention in that it synthesizes the same polyketide, 6-deoxyerythronolide B (6-dEB), synthesized by the megalomicin PKS. In modular PKS enzymes such as DEBS and the megalomicin PKS, the enzymatic steps for each round of condensation and reduction are encoded within a single "module" of the polypeptide (i.e., one distinct module for every condensation cycle). DEBS consists of a loading module and 6 extender modules and a chain terminating thioesterase (TE) domain within three extremely large polypeptides encoded by three open reading frames (ORFs, designated *eryAI*, *eryAII*, and *eryAIII*).

Each of the three polypeptide subunits of DEBS (DEBSI, DEBSII, and DEBSIII) contains 2 extender modules, DEBSI additionally contains the loading module. Collectively, these proteins catalyze the condensation and appropriate reduction of 1 propionyl CoA starter unit and 6 methylmalonyl CoA extender  
5 units. Modules 1, 2, 5, and 6 contain KR domains; module 4 contains a complete set, KR/DH/ER, of reductive and dehydratase domains; and module 3 contains no functional reductive domain. Following the condensation and appropriate dehydration and reduction reactions, the enzyme bound intermediate is lactonized by the TE at the end of extender module 6 to form 6-dEB.

10 More particularly, the loading module of DEBS consists of two domains, an acyl-transferase (AT) domain and an acyl carrier protein (ACP) domain. In other PKS enzymes, the loading module is not composed of an AT and an ACP but instead utilizes an inactivated KS, an AT, and an ACP. This inactivated KS is in most instances called KS<sup>Q</sup>, where the superscript letter is the abbreviation for  
15 the amino acid, glutamine, that is present instead of the active site cysteine required for activity. The AT domain of the loading module recognizes a particular acyl-CoA (propionyl for DEBS, which can also accept acetyl) and transfers it as a thiol ester to the ACP of the loading module. Concurrently, the AT on each of the extender modules recognizes a particular extender-CoA  
20 (methylmalonyl for DEBS) and transfers it to the ACP of that module to form a thioester. Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module migrates to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module 1 possesses an acyl-KS and a methylmalonyl ACP. The acyl group derived from the loading module is  
25 then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading unit (elongation or extension). The growing polyketide chain is transferred from the ACP to the KS of the next module, and the process continues.

30 The polyketide chain, growing by two carbons each module, is sequentially passed as a covalently bound thiol ester from module to module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the

name polyketide arises. Commonly, however, the beta keto group of each two-carbon unit is modified just after it has been added to the growing polyketide chain but before it is transferred to the next module by either a KR, a KR plus a DH, or a KR, a DH, and an ER. As noted above, modules may contain additional enzymatic activities as well.

Once a polyketide chain traverses the final extender module of a PKS, it encounters the releasing domain or thioesterase found at the carboxyl end of most PKSs. Here, the polyketide is cleaved from the enzyme and cyclized. The resulting polyketide can be modified further by tailoring or modification enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, the final steps in conversion of 6-dEB to erythromycin A include the actions of a number of modification enzymes, such as: C-6 hydroxylation, attachment of mycarose and desosamine sugars, C-12 hydroxylation (which produces erythromycin C), and conversion of mycarose to cladinose via *O*-methylation, as shown in Figure 5.

With this overview of PKS and post-PKS modification enzymes, one can better appreciate the recombinant megalomicin biosynthetic genes provided by the invention and their function, as described in the following Section.

20

### Section III: The Megalomicin Biosynthetic Genes and Nucleic Acid Fragments

The megalomicin PKS was isolated and cloned by the following procedure. Genomic DNA was isolated from a megalomicin producing strain of *Micromonospora megalomicea* subsp. *nigra* (ATCC 27598), partially digested with a restriction enzyme, and cloned into a commercially available cosmid vector to produce a genomic library. This library was then probed with probe generated from the erythromycin biosynthetic genes as well as from cosmids identified as containing sequences homologous to erythromycin biosynthetic genes. This probing identified a set of cosmids, which were analyzed by DNA sequence analysis and restriction enzyme digestion, which revealed that the desired DNA had been isolated and that the entire PKS gene cluster was contained in overlapping segments on four of the cosmids identified. Figure 1 shows the cosmids, and the portions of the megalomicin biosynthetic gene cluster in the

30



insert DNA of the cosmids. Figure 1 shows that the complete megalomicin biosynthetic gene cluster is contained within the insert DNA of cosmids pKOS079-138B, pKOS079-124B, pKOS079-93D, and pKOS079-93A. Each of these cosmids has been deposited with the American Type Culture Collection in accordance with the terms of the Budapest Treaty (cosmid pKOS079-138B is available under accession no. ATCC \_\_\_\_\_; cosmid pKOS079-124B is available under accession no. ATCC \_\_\_\_\_; cosmid pKOS079-93D is available under accession no. ATCC \_\_\_\_\_; and cosmid pKOS079-93A is available under accession no. ATCC \_\_\_\_\_). Various additional reagents of the invention can be isolated from these cosmids. DNA sequence analysis was also performed on the various subclones of the invention, as described herein. Further analysis of these cosmids and subclones prepared from the cosmids facilitated the identification of the location of various megalomicin biosynthetic genes, including the ORFs encoding the PKS, modules encoded by those ORFs, and coding sequences for megalomicin modification enzymes. The location of these genes and modules is shown on Figure 2.

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the megalomicin PKS and other biosynthetic enzymes and other biosynthetic enzymes of *Micromonospora megalomicea* is shown herein merely to illustrate a preferred embodiment of the invention, and the invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences encoded by the DNA sequences shown herein merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following description of the various regions of the megalomicin PKS and the megalomicin modification

enzymes and corresponding coding sequences is provided. To facilitate description of the invention, reference to a PKS, protein, module, or domain herein can also refer to DNA compounds comprising coding sequences therefor and *vice versa*. Also, unless otherwise indicated, reference to a heterologous PKS refers to a PKS or DNA compounds comprising coding sequences therefor from an organism other than *Micromonospora megalomicea*. In addition, reference to a PKS or its coding sequence includes reference to any portion thereof.

Thus, the invention provides DNA molecules in isolated (i.e., not pure, but existing in a preparation in an abundance and/or concentration not found in nature) and purified (i.e., substantially free of contaminating materials or substantially free of materials with which the corresponding DNA would be found in nature) form. The DNA molecules of the invention comprise one or more sequences that encode one or more domains (or fragments of such domains) of one or more modules in one or more of the ORFs of the megalomicin PKS and sequences that encode megalomicin modification enzymes from the megalomicin biosynthetic gene cluster. Examples of PKS domains include the KS, AT, DH, KR, ER, ACP, and TE domains of at least one of the 6 extender modules and loading module of the three proteins encoded by the three ORFs of the megalomicin PKS gene cluster. Examples of megalomicin modification enzymes include those that synthesize the mycarose, desosamine, and megosamine moieties, those that transfer those sugar moieties to the polyketide 6-dEB, those that hydroxylate the polyketide at C-6 and C-12, and those that acylate the sugar moieties.

In an especially preferred embodiment, the DNA molecule is a recombinant DNA expression vector or plasmid, as described in more detail in the following Section. Generally, such vectors can either replicate in the cytoplasm of the host cell or integrate into the chromosomal DNA of the host cell. In either case, the vector can be a stable vector (i.e., the vector remains present over many cell divisions, even if only with selective pressure) or a transient vector (i.e., the vector is gradually lost by host cells with increasing numbers of cell divisions).

The megalomicin PKS gene cluster comprises three ORFs (*megAI*, *megAII*, and *megAIII*). Each ORF encodes two extender modules of the PKS; the first ORF also encodes the loading module. Each extender module is composed of at least a KS, an AT, and an ACP domain. The locations of the various encoding regions of

these ORFs are shown in Figure 2 and described with reference to the sequence information below. The megalomicin PKS produces the polyketide known as 6-dEB, shown in Figure 4. In megalomicin-producing organisms, 6-dEB is converted to erythromycin C by a set of modification enzymes. Thus, 6-dEB is  
5 converted to erythronolide B by the *megF* gene product (a homolog of the *eryF* gene product), then to 3- $\alpha$ -mycarosyl-erythronolide B by the *megBV* gene product (a homolog of the *eryBV* gene product), then to erythromycin D by the *megCIII* gene product (a homolog of the *eryCIII* gene product), then to erythromycin C by the *megK* gene product (a homolog of the *eryK* gene product).

10 In addition to these modification enzymes, such megalomicin-producing organisms also contain the modification enzymes necessary for the biosynthesis of the desosamine and mycarose moieties that are similarly utilized in erythromycin biosynthesis, as shown in Figure 5. Megalomicin A contains the complete erythromycin C structure, and its biosynthesis additionally involves the formation  
15 of L-megosamine (L-rhodamine) and its attachment to the C-6 hydroxyl (Figures 3 and 5, inset), followed by acylation of the C-3''' and(or) C-4''' hydroxyls as the terminal steps. L-megosamine is the same as *N*-dimethyl-L-daunosamine; the daunosamine genes have been characterized from *Streptomyces peucetius* (see Colombo and Hutchinson, *J. Indust. Microbiol. Biotechnol.*, in  
20 press; Otten *et al.*, 1996, *J Bacteriol* 178:7316-7321, and references cited therein). Some of the rhodamine genes also have been cloned and partially characterized from another anthracycline producing *Streptomyces* sp. (see Torkkell *et al.*, 1997, *Mol. Gen. Genet.* 256(2):203-209). Because the timing of the glycosylation with TDP-megosamine in relation to the addition of mycarose and desosamine to  
25 erythronolide B, plus the C-12 hydroxylation, is unknown, the pathway could involve a different order of glycosylation and C-12 hydroxylation steps than the one shown in Figure 5. Regardless, the megalomicin biosynthetic gene cluster contains the genes to make L-rhodamine and attach it to the correct macrolide substrate.

30 The biosynthetic pathways to make the glycosides desosamine, mycarose, and megosamine are shown in Figure 6. The present invention provides the genes for each biosynthetic pathway shown in this Figure, and these recombinant genetic

pathways can be used alone or in any combination to confer the pathway to a heterologous host.

The megalomicin PKS locus is similar to the *eryA* locus in size and organization. Most of the deoxysugar biosynthesis genes are homologs of the *eryB* mycarose and *eryC* desosamine biosynthesis and glycosyl attachment genes from *Saccharopolyspora erythraea* (see Summers *et al.*, 1997, *Microbiol. 143*:3251-3262; Haydock *et al.*, 1991, *Mol. Gen. Genet.* 230:120-128; Gaisser *et al.*, 1997, *Mol Gen Genet*, 256:239-251; Gaisser *et al.*, 1998, *Mol Gen Genet.* 257:78-88, incorporated herein by reference) or the *picC* homologs from the picromycin and narbomycin producer (see PCT patent publication No. 99/61599 and Xue *et al.*, 1998, *Proc. Nat. Acad. Sci. USA* 95, 12111-12116, incorporated herein by reference). The TDP-megosamine biosynthesis genes are homologs of the *dnm* genes (see Figure 5) and the pikromycin N-dimethyltransferase gene or its homologs reported in a cluster of L-rhodamine biosynthesis genes. The putative TDP-megosamine glycosyltransferase gene product (*geneX* in Figure 5) closely resembles the deduced products of the *eryBV*, *eryCIII*, *dnmS*, and pikromycin *desVII* genes, even though it recognizes different substrates than the products of each of these genes.

The following Table 1 shows the location of the genes in the *Micromonospora megalomicea* megalomicin biosynthetic pathway in the DNA sequence set forth in SEQ ID NO:1 (see also Figure 7; note some gene designations maybe different in Figure 7).

Table 1. Megalomicin Biosynthetic Gene Cluster  
*Micromonospora megalomicea* subsp. *nigra* (ATCC27598)

Location	Description
1..2451	sequence from cosmid pKOS079-138B
complement(1..144)	<i>megBVI</i> (or <i>megT</i> ), TDP-4-keto-6-deoxyglucose-
30 2,3-dehydratase	<i>megDVI</i> , TDP-4-keto-6-deoxyglucose 3,4-isomerase
928..2061	<i>megDI</i> , TDP-megosaminyl transferase ( <i>eryCIII</i>
2072..3382	homolog)
2452..40397	sequence of cosmid pKOS079-93D
35 3462..4634	<i>megG</i> (or <i>megY</i> ), mycarosyl acyltransferase
4651..5775	<i>megDII</i> , deoxysugar transaminase ( <i>eryCI</i> , <i>DnrJ</i>
	homolog)

	5822..6595	<i>megDIII</i> , TDP-daunosaminyl-N,N-
	dimethyltransferase	( <i>eryCVI</i> homolog)
5	6592..7197	<i>megDIV</i> , TDP-4-keto-6-deoxyglucose 3,5-epimerase
		( <i>eryBVII</i> , <i>dnmU</i> homolog)
	7220..8206	<i>megDV</i> , TDP-hexose 4-ketoreductase ( <i>eryBIV</i> ,
	<i>dnmV</i>	homolog)
10	complement(8228..9220)	<i>megBII-1</i> or <i>megDVII</i> , TDP-4-keto-L-6-deoxy-
	hexose 2,3-reductase	
	complement(9226..10479)	<i>megBV</i> , TDP-mycarosyl transferase
	complement(10483..11424)	<i>megBIV</i> , TDP-hexose 4-ketoreductase
	12181..22821	<i>megAI</i>
	12181..13791	Loading Module (L)
15	12505..13470	AT-L
	13576..13791	ACP-L
	13849..18207	Extender Module 1 (1)
	13849..15126	KS1
	15427..16476	AT1
20	17155..17694	KR1
	17947..18207	ACP1
	18268..22575	Extender Module 2 (2)
	18268..19548	KS2
	19876..20910	AT2
25	21517..22053	KR2
	22318..22575	ACP2
	22867..33555	<i>megAII</i>
	22957..27258	Extender Module 3 (3)
	22957..24237	KS3
30	24544..25581	AT3
	26230..26733	KR3 (inactive)
	26998..27258	ACP3
	27313..33312	Extender Module 4 (4)
	27393..28590	KS4
35	28897..29931	AT4
	29953..30477	DH4
	31396..32244	ER4
	32257..32799	KR4
	33052..33312	ACP4
40	33666..43271	<i>megAIII</i>
	33780..38120	Extender Module 5 (5)
	33780..35027	KS5
	35385..36419	AT5
	37068..37604	KR5
45	37860..38120	ACP5
	38187..42425	Extender Module 6 (6)
	38187..39470	KS6
	39795..40811	AT6
	40398..46641	sequences from cosmid pKOS079-93A

	41406..41936	KR6
	42168..42425	ACP6
	42585..43271	TE
	43268..44344	<i>megCII</i> , TDP-4-keto-6-deoxyglucose 3,4-isomerase
5	44355..45623	<i>megCIII</i> , TDP-desosaminy transferase
	45620..46591	<i>megBII</i> , TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase
	complement(46660..47403)	<i>megH</i> , TEII
	complement(47411..47980)	<i>megF</i> , C-6 hydroxylase

10

In a specific embodiment, the invention provides an isolated nucleic acid fragment comprising a nucleotide sequence encoding a domain of the megalomicin polyketide synthase or a megalomicin modification enzyme. The isolated nucleic acid fragment can be a DNA or a RNA. Preferably, the isolated nucleic acid fragment is a recombinant DNA compound. A nucleotide sequence that is complementary to the nucleotide sequence encoding a domain of megalomicin PKS or a megalomicin modification enzyme is also provided.

The isolated nucleic acid fragment can comprise a single, multiple or all the open reading frame(s) (ORF) of the megalomicin PKS or the megalomicin modification enzyme. Exemplary ORFs of megalomicin PKS include the ORFs of the *megAI*, *megAII* and *megAIII* genes. The isolated nucleic acids of the invention also include nucleic acids that encode one or more domains and one or more modules of the megalomicin PKS. Exemplary domains of the megalomicin PKS include a TE domain, a KS domain, an AT domain, an ACP domain, a KR domain, a DH domain and an ER domain. In a preferred embodiment, the nucleic acid comprises the coding sequence for a loading module, a thioesterase domain, and all six extender modules of the megalomicin PKS.

Megalomicin modification enzymes include those enzymes involved in the conversion of 6-DEB into a megalomicin such as the enzymes encoded by *megF*, *megBV*, *megCIII*, *megK*, *megDI* and *megG* (or *megY*). Megalomicin modification enzymes also include those enzymes involved in the biosynthesis of mycarose, megosamine or desosamine, which are used as biosynthetic intermediates in the biosynthesis of various megalomicin species and other related polyketides. The enzymes that are involved in biosynthesis of mycarose, megosamine or desosamine are described in Figures 5 and 10. The megalomicin PKS and megalomicin modification enzymes are collectively referred to as megalomicin



biosynthetic enzymes; the genes encoding such enzymes are collectively referred to as megalomicin biosynthetic genes; and nucleic acids that comprise a portion of or entire megalomicin biosynthetic genes are collectively referred to as megalomicin biosynthetic nucleic acid(s).

5 In specific embodiments, the megalomicin biosynthetic nucleic acids comprise the sequence of SEQ ID NO:1, or the coding regions thereof, or nucleotide sequences encoding, in whole or in part, a megalomicin biosynthetic enzyme protein. The isolated nucleic acids typically consists of at least 25 (continuous) nucleotides, 50 nucleotides, 100 nucleotides, 150 nucleotides, or 200  
10 nucleotides of megalomicin biosynthetic nucleic acid sequence, or a full-length megalomicin biosynthetic coding sequence. In another embodiment, the nucleic acids are smaller than 35, 200, or 500 nucleotides in length. Nucleic acids can be single or double stranded. Nucleic acids that hybridize to or are complementary to the foregoing sequences, in particular the inverse complement to nucleic acids that  
15 hybridize to the foregoing sequences (*i.e.*, the inverse complement of a nucleic acid strand has the complementary sequence running in reverse orientation to the strand so that the inverse complement would hybridize without mismatches to the nucleic acid strand) are also provided. In specific aspects, nucleic acids are provided which comprise a sequence complementary to (specifically are the  
20 inverse complement of) at least 10, 25, 50, 100, or 200 nucleotides or the entire coding region of a megalomicin biosynthetic gene.

The megalomicin biosynthetic nucleic acids provided herein include those with nucleotide sequences encoding substantially the same amino acid sequences as found in native megalomicin biosynthetic enzyme proteins, and those encoding  
25 amino acid sequences with functionally equivalent amino acids, as well as megalomicin biosynthetic enzyme derivatives or analogs as described in Section IV.

Some regions within the megalomicin PKS genes are highly homologous or identical to one another, as can be readily identified by an analysis of the  
30 sequence. The coding sequence for the KS and AT domains of module 2 shares significant identity with the coding sequence for the KS and AT domains of module 6. This sequence homology or identity at the nucleic acid, *e.g.*, DNA, level can render the nucleic acid unstable in certain host cells. To improve the stability

of the nucleic acids comprising a portion or the entire megalomicin PKS genes and megalomicin modification enzyme genes, the nucleic acid or DNA sequences can be changed to reduce or abolish the sequence homology or identity. Preferably, the DNA codons of homologous regions within the PKS or the megalomicin  
5 modification enzyme coding sequence are changed to reduce or abolish the sequence homology or identity without changing the amino acid sequences encoded by said changed DNA codons (see the examples below). The stability of the nucleic acid or DNA can also be improved by codon changes that reduce or abolish the sequence homology or identity while also changing the amino acid  
10 sequence, provided that the amino acid sequence change(s) does not substantially change the desired activity of the encoded megalomicin PKS. Thus, for example, one can simply substitute for the *megAIII* ORF an ORF from *eryAIII*, *oleAIII*, *picAIII*, or *picAIV* genes.

The recombinant DNA compounds of the invention that encode the  
15 megalomicin PKS and modification proteins or portions thereof are useful in a variety of applications. While many of these applications relate to the heterologous expression of the megalomicin biosynthetic genes or the construction of hybrid PKS enzymes, many useful applications involve the natural megalomicin producer *Micromonospora megalomicea*. For example, one can use the recombinant DNA  
20 compounds of the invention to disrupt the megalomicin biosynthetic genes by homologous recombination in *Micromonospora megalomicea*. The resulting host cell is a preferred host cell for making polyketides modified by oxidation, hydroxylation, glycosylation, and acylation in a manner similar to megalomicin, because the genes that encode the proteins that perform these reactions are of  
25 course present in the host cell, and because the host cell does not produce megalomicin that could interfere with production or purification of the polyketide of interest.

One illustrative recombinant host cell provided by the present invention expresses a recombinant megalomicin PKS in which the module 1 KS domain is  
30 inactivated by deletion or other mutation. In a preferred embodiment, the inactivation is mediated by a change in the KS domain that renders it incapable of binding substrate (called a KS1° mutation). In a particularly preferred embodiment, this inactivation is rendered by a mutation in the codon for the active

site cysteine that changes the codon to another codon, such as an alanine codon. Such constructs are especially useful when placed in translational reading frame with extender modules 1 and 2 of a megalomicin or the corresponding modules of another PKS. The utility of these constructs is that host cells expressing, or cell  
5 free extracts containing, a PKS comprising the protein encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare a polyketide of interest. See U.S. patent application Serial No. 09/492,773, filed 27 Jan. 2000, and PCT patent publication No. 00/44717, both of which are incorporated herein by reference. Such KS1° constructs of the invention are useful  
10 in the production of 13-substituted-megalomicin compounds in *Micromonospora megalomicea* host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl.

In a variant of this embodiment, one can employ a megalomicin PKS in  
15 which the ACP domain of module 1 has been rendered inactive. In another embodiment, one can delete the loading domain of the megalomicin PKS and provide monoketide substrates for processing by the remainder of the PKS.

The compounds of the invention can also be used to construct recombinant host cells of the invention in which coding sequences for one or more domains or  
20 modules of the megalomicin PKS or for another megalomicin biosynthetic gene have been deleted by homologous recombination with the *Micromonospora megalomicea* chromosomal DNA. Those of skill in the art will appreciate that the compounds used in the recombination process are characterized by their homology with the chromosomal DNA and not by encoding a functional protein due to their  
25 intended function of deleting or otherwise altering portions of chromosomal DNA. For this and a variety of other applications, the compounds of the present invention include not only those DNA compounds that encode functional proteins but also those DNA compounds that are complementary or identical to any portion of the megalomicin biosynthetic genes.

30 Thus, the invention provides a variety of modified *Micromonospora megalomicea* host cells in which one or more of the megalomicin biosynthetic genes have been mutated or disrupted. Transformation systems for *M. megalomicea* have been described by Hasegawa *et al.*, 1991, *J. Bacteriol.*

173:7004-11; and Takada *et al.*, 1994, *J. Antibiot.* 47:1167-1170, both of which are incorporated herein by reference. These cells are especially useful when it is desired to replace the disrupted function with a gene product expressed by a recombinant DNA expression vector. While such expression vectors of the invention are described in more detail in the following Section, those of skill in the art will appreciate that the vectors have application to *M. megalomicea* as well. Such *M. megalomicea* host cells can be preferred host cells for expressing megalomicin derivatives of the invention. Particularly preferred host cells of this type include those in which the coding sequence for the loading module has been mutated or disrupted, those in which one or more of any of the PKS gene ORFs has been mutated or disrupted, and/or those in which the genes for one or more modification (glycosylation, acylation, hydroxylation) have been mutated or disrupted.

While the present invention provides many useful compounds having application to, and recombinant host cells derived from, *Micromonospora megalomicea*, many important applications of the present invention relate to the heterologous expression of all or a portion of the megalomicin biosynthetic genes in cells other than *M. megalomicea*, as described in Section V.

#### 20 Section IV: The Megalomicin Biosynthetic Enzymes and Antibodies Recognizing such Enzymes

In another specific embodiment, the invention provides a substantially purified polypeptide, which is encoded by a nucleic acid fragment comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. The polypeptide can comprise a single domain, multiple domains or a full-length megalomicin PKS or megalomicin modification enzyme. Functional fragments, analogs or derivatives of the megalomicin PKS or megalomicin modification enzyme polypeptides are also provided. Preferably, such fragments, analogs or derivatives can be recognized an antibody raised against a megalomicin PKS or megalomicin modification enzyme. Also preferably, such fragments, analogs or derivatives comprise an amino acid sequence that has at least 60% identity, more preferably at least 90% identity to their wild type counterparts.

An exemplary nucleotide sequence encoding, and the corresponding amino acid sequence of, a megalomicin biosynthetic enzyme is disclosed in SEQ ID NO:1. Homologs (*e.g.*, nucleic acids of the above-listed genes of species other than *Micromonospora megalomicea*) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular sequence provided as a probe using methods well known in the art for nucleic acid hybridization and cloning (*e.g.*, as described in Section III) in accordance with the methods of the present invention.

The megalomicin biosynthetic enzyme proteins, or domains thereof, of the present invention can be obtained by methods well known in the art for protein purification and recombinant protein expression in accordance with the methods of the present invention. For recombinant expression of one or more of the proteins, the nucleic acid containing all or a portion of the nucleotide sequence encoding the protein can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein coding sequence. Transcriptional and translational signals can be supplied by the native promoter for a megalomicin biosynthetic gene and/or flanking regions.

A variety of host-vector systems may be utilized to express the protein coding sequence. These include but are not limited to mammalian cell systems infected with virus (*e.g.* vaccinia virus, adenovirus, and the like); insect cell systems infected with virus (*e.g.* baculovirus); microorganisms such as yeast containing yeast vectors; or bacteria transformed with bacteriophage, DNA, plasmid DNA, or cosmid DNA. The expression elements of vectors vary in their properties. Depending on the host-vector system utilized, any one of a number of suitable transcription and translation elements may be used.

In a specific embodiment, a vector is used that comprises a promoter operably linked to nucleic acid sequences encoding a megalomicin biosynthetic enzyme, or a domain, fragment, derivative or homolog, thereof, one or more origins of replication, and optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene).

Expression vectors containing the sequences of interest can be identified by three general approaches: (a) nucleic acid hybridization, (b) presence or

absence of "marker" gene function, and (c) expression of the inserted sequences. In the first approach, megalomicin biosynthetic nucleic acid sequences can be detected by nucleic acid hybridization to probes comprising sequences homologous and complementary to the inserted sequences. In the second  
5 approach, the recombinant vector/host system can be identified and selected based upon the presence or absence of certain "marker" functions (*e.g.*, binding to an anti-megalomicin biosynthetic enzyme antibody, resistance to antibiotics, occlusion body formation in baculovirus, and the like) caused by insertion of the sequences of interest in the vector. For example, if a megalomicin biosynthetic  
10 gene, or portion thereof, is inserted within the marker gene sequence of the vector, recombinants containing the megalomicin biosynthetic gene fragment will be identified by the absence of the marker gene function. In the third approach, recombinant expression vectors can be identified by assaying for the megalomicin biosynthetic gene products expressed by the recombinant vector. Such assays can  
15 be based, for example, on the physical or functional properties of the interacting species in *in vitro* assay systems, *e.g.*, megalomicin synthesis activity, immunoreactivity to antibodies specific for the protein.

Once recombinant megalomicin biosynthetic genes or nucleic acids are identified, several methods known in the art can be used to propagate them in  
20 accordance with the methods of the present invention. Once a suitable host system and growth conditions have been established, recombinant expression vectors can be propagated and amplified in quantity. As previously described, the expression vectors or derivatives which can be used include, but are not limited to: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such  
25 as baculovirus, yeast vectors; bacteriophage vectors such as lambda phage; and plasmid and cosmid vectors.

In addition, a host cell strain may be chosen that modulates the expression of the inserted sequences, or modifies or processes the expressed proteins in the specific fashion desired. Expression from certain promoters can be elevated in the  
30 presence of certain inducers; thus expression of the genetically-engineered megalomicin biosynthetic enzymes may be controlled. Furthermore, different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (*e.g.* glycosylation, phosphorylation, and



the like) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein is achieved. For example, expression in a bacterial system can be used to produce an unglycosylated core protein, while expression in mammalian cells ensures  
5 "native" glycosylation of a heterologous protein. Furthermore, different vector/host expression systems may effect processing reactions to different extent.

In particular, megalomicin biosynthetic enzyme derivatives can be made by altering their sequences by substitutions, additions or deletions that provide for functionally equivalent molecules. Due to the degeneracy of nucleotide coding  
10 sequences, other DNA sequences which encode substantially the same amino acid sequence as an megalomicin biosynthetic gene can be used in the practice of the present invention. These include but are not limited to nucleotide sequences comprising all or portions of megalomicin biosynthetic genes that are altered by the substitution of different codons that encode the amino acid residue within the  
15 sequence, thus producing a silent change. Likewise, the megalomicin biosynthetic enzyme derivatives of the invention include, but are not limited to, those containing, as a primary amino acid sequence, all or part of the amino acid sequence of megalomicin biosynthetic enzymes, including altered sequences in which functionally equivalent amino acid residues are substituted for residues  
20 within the sequence resulting in a silent change. For example, one or more amino acid residues within the sequence can be substituted by another amino acid of a similar polarity which acts as a functional equivalent, resulting in a silent alteration. Substitutes for an amino acid within the sequence may be selected from other members of the class to which the amino acid belongs. For example,  
25 the nonpolar (hydrophobic) amino acids include alanine, leucine, isoleucine, valine, proline, phenylalanine, tryptophan and methionine. The polar neutral amino acids include glycine, serine, threonine, cysteine, tyrosine, asparagine, and glutamine. The positively charged (basic) amino acids include arginine, lysine and histidine. The negatively charged (acidic) amino acids include aspartic acid and  
30 glutamic acid.

In a specific embodiment of the invention, the nucleic acids encoding proteins and proteins consisting of or comprising a domain or a fragment of megalomicin biosynthetic enzyme consisting of at least 6 (continuous) amino

acids are provided. In other embodiments, the domain or fragment consists of at least 10, 20, 30, 40, or 50 amino acids of a megalomicin biosynthetic enzyme. In specific embodiments, such domains or fragments are not larger than 35, 100 or 200 amino acids. Derivatives or analogs of megalomicin biosynthetic enzyme  
5 include but are not limited to molecules comprising regions that are substantially homologous to megalomicin biosynthetic enzyme in various embodiments, at least 30%, 40%, 50%, 60%, 70%, 80%, 90% or 95% identity over an amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art in  
10 accordance with the methods of the present invention or whose encoding nucleic acid is capable of hybridizing to a sequence encoding a megalomicin biosynthetic enzyme under stringent, moderately stringent, or nonstringent conditions.

The megalomicin biosynthetic enzyme domains, derivatives and analogs of the invention can be produced by various methods known in the art in accordance  
15 with the methods of the present invention. The manipulations which result in their production can occur at the gene or protein level. For example, the cloned megalomicin biosynthetic gene sequence can be modified by any of numerous strategies known in the art (Sambrook et al., 1990, *Molecular Cloning, A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor,  
20 New York) in accordance with the methods of the present invention. The sequences can be cleaved at appropriate sites with restriction endonuclease(s), followed by further enzymatic modification if desired, isolated, and ligated *in vitro*.

Additionally, the megalomicin biosynthetic enzyme-encoding nucleotide  
25 sequence can be mutated *in vitro* or *in vivo*, to create and/or destroy translation, initiation, and/or termination sequences, or to create variations in coding regions and/or form new restriction endonuclease sites or destroy pre-existing ones, to facilitate further *in vitro* modification. Any technique for mutagenesis known in the art can be used in accordance with the methods of the present invention,  
30 including but not limited to, chemical mutagenesis and *in vitro* site-directed mutagenesis (Hutchinson et al., *J. Biol. Chem.* 253:6551-6558 (1978)), use of TAB® linkers (Pharmacia), and the like.

Once a recombinant cell expressing a megalomicin biosynthetic enzyme protein, or a domain, fragment or derivative thereof, is identified, the individual gene product can be isolated and analyzed. This is achieved by assays based on the physical and/or functional properties of the protein, including, but not limited to, radioactive labeling of the product followed by analysis by gel electrophoresis, immunoassay, cross-linking to marker-labeled product, and the like.

The megalomicin biosynthetic enzyme proteins may be isolated and purified by standard methods known in the art or recombinant host cells expressing the complexes or proteins in accordance with the methods of the invention, including but not restricted to column chromatography (*e.g.*, ion exchange, affinity, gel exclusion, reversed-phase high pressure, fast protein liquid, and the like), differential centrifugation, differential solubility, or by any other standard technique used for the purification of proteins. Functional properties may be evaluated using any suitable assay known in the art in accordance with the methods of the present invention.

Alternatively, once a megalomicin biosynthetic enzyme or its domain or derivative is identified, the amino acid sequence of the protein can be deduced from the nucleotide sequence of the gene which encodes it. As a result, the protein or its domain or derivative can be synthesized by standard chemical methods known in the art in accordance with the methods of the present invention (see Hunkapiller et al, *Nature* 310:105-111 (1984)).

Manipulations of megalomicin biosynthetic enzymes may be made at the protein level. Included within the scope of the invention are megalomicin biosynthetic enzyme domains, derivatives or analogs or fragments, which are differentially modified during or after translation, *e.g.*, by glycosylation, acetylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to an antibody molecule or other cellular ligand, and the like. Any of numerous chemical modifications may be carried out by known techniques, including but not limited to specific chemical cleavage by cyanogen bromide, trypsin, chymotrypsin, papain, V8 protease, NaBH<sub>4</sub>, acetylation, formylation, oxidation, reduction, metabolic synthesis in the presence of tunicamycin, and the like.

In specific embodiments, the megalomicin biosynthetic enzymes are modified to include a fluorescent label. In other specific embodiments, the megalomicin biosynthetic enzyme is modified to have a heterofunctional reagent, such heterofunctional reagents can be used to crosslink the members of the  
5 complex.

In addition, domains, analogs and derivatives of a megalomicin biosynthetic enzyme can be chemically synthesized. For example, a peptide corresponding to a portion of a megalomicin biosynthetic enzyme, which comprises the desired domain or which mediates the desired activity *in vitro* can  
10 be synthesized by use of a peptide synthesizer. Furthermore, if desired, nonclassical amino acids or chemical amino acid analogs can be introduced as a substitution or addition into the megalomicin biosynthetic enzyme sequence. Non-classical amino acids include but are not limited to the D-isomers of the common amino acids, alpha-amino isobutyric acid, 4-aminobutyric acid,  
15 2-aminobutyric acid, 6-amino hexanoic acid, Aib, 2-amino isobutyric acid, 3-amino propionic acid, ornithine, norleucine, norvaline, hydroxyproline, sarcosine, citrulline, cysteic acid, t-butylglycine, t-butylalanine, phenylglycine, cyclohexylalanine,  $\beta$ -alanine, fluoro-amino acids, designer amino acids such as  $\beta$ -methyl amino acids, Ca-methyl amino acids, Na-methyl amino acids, and amino  
20 acid analogs in general. Furthermore, the amino acid can be D (dextrorotary) or L (levorotary).

In cases where natural products are suspected of being mutant or are isolated from new species, the amino acid sequence of the megalomicin biosynthetic enzyme isolated from the natural source, as well as those expressed *in*  
25 *vitro*, or from synthesized expression vectors *in vivo* or *in vitro*, can be determined from analysis of the DNA sequence, or alternatively, by direct sequencing of the isolated protein. Such analysis may be performed by manual sequencing or through use of an automated amino acid sequenator.

The megalomicin biosynthetic enzyme proteins may also be analyzed by  
30 hydrophilicity analysis (Hopp and Woods, *Proc. Natl. Acad. Sci. USA* 78:3824-3828 (1981)). A hydrophilicity profile can be used to identify the hydrophobic and hydrophilic regions of the proteins, and help predict their orientation in designing substrates for experimental manipulation, such as in binding

experiments, antibody synthesis, and the like. Secondary structural analysis can also be done to identify regions of the megalomicin biosynthetic enzyme that assume specific structures (Chou and Fasman, *Biochemistry* 13:222-23 (1974)). Manipulation, translation, secondary structure prediction, hydrophilicity and  
5 hydrophobicity profiles, open reading frame prediction and plotting, and determination of sequence homologies, can be accomplished using computer software programs available in the art.

Other methods of structural analysis including but not limited to X-ray crystallography (Engstrom, *Biochem. Exp. Biol.* 11:7-13 (1974)), mass  
10 spectroscopy and gas chromatography (Methods in Protein Science, J. Wiley and Sons, New York, 1997), and computer modeling (Fletteriek and Zoller, eds., 1986, Computer Graphics and Molecular Modeling, In: *Current Communications in Molecular Biology*, Cold Spring Harbor Laboratory, Cold Spring Harbor Press, New York) can also be employed.

15 The invention also provides an antibody, or a fragment or derivative thereof, which immuno-specifically binds to a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme. In a specific embodiment, an antibody which immuno-specifically binds to a domain of the megalomicin biosynthetic enzyme encoded by a nucleic acid that hybridizes to a  
20 nucleic acid having the nucleotide sequence set forth in the SEQ. ID NO:1, or a fragment or derivative of said antibody containing the binding domain thereof is provided. Preferably, the antibody is a monoclonal antibody.

The megalomicin biosynthetic enzyme protein and domains, fragments, homologs and derivatives thereof may be used as immunogens to generate  
25 antibodies which immunospecifically bind such immunogens. Such antibodies include but are not limited to polyclonal, monoclonal, chimeric, single chain, Fab fragments, and an Fab expression library.

Various procedures known in the art may be used for the production of polyclonal antibodies to a megalomicin biosynthetic enzyme protein of the  
30 invention, its domains, derivatives, fragments or analogs in accordance with the methods of the present invention.

For production of the antibody, various host animals can be immunized by injection with the native megalomicin biosynthetic enzyme protein or a synthetic

version, or a derivative of the foregoing, such as a cross-linked megalomicin biosynthetic enzyme. Such host animals include but are not limited to rabbits, mice, rats, and the like. Various adjuvants can be used to increase the immunological response, depending on the host species, and include but are not limited to Freund's (complete and incomplete), mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, and potentially useful human adjuvants such as bacille Calmette-Guerin (BCG) and corynebacterium parvum.

For preparation of monoclonal antibodies directed towards a megalomicin biosynthetic enzyme or domains, derivatives, fragments or analogs thereof, any technique that provides for the production of antibody molecules by continuous cell lines in culture may be used. Such techniques include but are not restricted to the hybridoma technique originally developed by Kohler and Milstein (*Nature* 256:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., *Immunology Today* 4:72 (1983)), and the EBV hybridoma technique to produce human monoclonal antibodies (Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). In an additional embodiment, monoclonal antibodies can be produced in germ-free animals (WO89/12690). Human antibodies may be used and can be obtained by using human hybridomas (Cote et al., *Proc. Natl. Acad. Sci. USA* 80:2026-2030 (1983)) or by transforming human B cells with EBV virus *in vitro* (Cole et al., in *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96 (1985)). Techniques developed for the production of "chimeric antibodies" (Morrison et al., *Proc. Natl. Acad. Sci. USA* 81:6851-6855 (1984); Neuberger et al., *Nature* 312:604-608 (1984); Takeda et al., *Nature* 314:452-454 (1985)) by splicing the genes from a mouse antibody molecule specific for the megalomicin biosynthetic enzyme protein together with genes from a human antibody molecule of appropriate biological activity can be used; such antibodies are within the scope of this invention.

Techniques described for the production of single chain antibodies (U.S. patent 4,946,778) can be adapted to produce megalomicin biosynthetic enzyme-specific single chain antibodies. An additional embodiment utilizes the techniques described for the construction of Fab expression libraries (Huse et al., *Science*



246:1275-1281 (1989)) to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity for megalomicin biosynthetic enzyme, or domains, derivatives, or analogs thereof. Non-human antibodies can be "humanized" by known methods (*see, e.g.*, U.S. Patent No. 5,225,539).

5       Antibody fragments that contain the idiotypes of a megalomicin biosynthetic enzyme can be generated by techniques known in the art in accordance with the methods of the present invention. For example, such fragments include but are not limited to: the F(ab')<sub>2</sub> fragment which can be produced by pepsin digestion of the antibody molecule; the Fab' fragments that  
10       can be generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragment, the Fab fragments that can be generated by treating the antibody molecular with papain and a reducing agent, and Fv fragments.

      In the production of antibodies, screening for the desired antibody can be accomplished by techniques known in the art in accordance with the methods of  
15       the present invention, *e.g.*, ELISA (enzyme-linked immunosorbent assay). To select antibodies specific to a particular domain of the megalomicin biosynthetic enzyme, one may assay generated hybridomas for a product that binds to the fragment of a megalomicin biosynthetic enzyme that contains such a domain.

      The foregoing antibodies can be used in methods known in the art relating  
20       to the localization and/or quantitation of megalomicin biosynthetic enzyme proteins, *e.g.*, for imaging these proteins or measuring levels thereof in samples, in accordance with the methods of the present invention.

#### Section V: Heterologous Expression of the Megalomicin Biosynthetic Genes

25       In one important embodiment, the invention provides methods for the heterologous expression of one or more of the megalomicin biosynthetic genes and recombinant DNA expression vectors useful in the method. For purposes of the invention, any host cell other than *Micromonospora megalomicea* is a heterologous host cell. Thus, included within the scope of the invention in  
30       addition to isolated nucleic acids encoding domains, modules, or proteins of the megalomicin PKS and modification enzymes, are recombinant expression vectors that include such nucleic acids. The term expression vector refers to a nucleic acid that can be introduced into a host cell or cell-free transcription and translation

system. An expression vector can be maintained permanently or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a promoter that drives expression of an RNA, which typically is translated into a polypeptide in the cell or cell extract. For efficient translation of RNA into protein, the expression vector also typically contains a ribosome-binding site sequence positioned upstream of the start codon of the coding sequence of the gene to be expressed. Other elements, such as enhancers, secretion signal sequences, transcription termination sequences, and one or more marker genes by which host cells containing the vector can be identified and/or selected, may also be present in an expression vector. Selectable markers, i.e., genes that confer antibiotic resistance or sensitivity, are preferred and confer a selectable phenotype on transformed cells when the cells are grown in an appropriate selective medium.

The various components of an expression vector can vary widely, depending on the intended use of the vector and the host cell(s) in which the vector is intended to replicate or drive expression. Expression vector components suitable for the expression of genes and maintenance of vectors in *E. coli*, yeast, *Streptomyces*, and other commonly used cells are widely known and commercially available. For example, suitable promoters for inclusion in the expression vectors of the invention include those that function in eucaryotic or procaryotic host cells. Promoters can comprise regulatory sequences that allow for regulation of expression relative to the growth of the host cell or that cause the expression of a gene to be turned on or off in response to a chemical or physical stimulus. For *E. coli* and certain other bacterial host cells, promoters derived from genes for biosynthetic enzymes, antibiotic-resistance conferring enzymes, and phage proteins can be used and include, for example, the galactose, lactose (*lac*), maltose, tryptophan (*trp*), beta-lactamase (*bla*), bacteriophage lambda PL, and T5 promoters. In addition, synthetic promoters, such as the *tac* promoter (U.S. Patent No. 4,551,433), can also be used.

Thus, recombinant expression vectors contain at least one expression system, which, in turn, is composed of at least a portion of the megalomicin PKS and/or other megalomicin biosynthetic gene coding sequences operably linked to a

promoter and optionally termination sequences that operate to effect expression of the coding sequence in compatible host cells. The host cells are modified by transformation with the recombinant DNA expression vectors of the invention to contain the expression system sequences either as extrachromosomal elements or  
5 integrated into the chromosome. The resulting host cells of the invention are useful in methods to produce PKS and post-PKS modification enzymes as well as polyketides and antibiotics and other useful compounds derived therefrom.

Preferred host cells for purposes of selecting vector components for expression vectors of the present invention include fungal host cells such as yeast  
10 and procaryotic host cells such as *E. coli* and *Streptomyces*, but mammalian host cells can also be used. In hosts such as yeasts, plants, or mammalian cells that ordinarily do not produce polyketides, it may be necessary to provide, also typically by recombinant means, suitable holo-ACP synthases to convert the recombinantly produced PKS to functionality. Provision of such enzymes is  
15 described, for example, in PCT publication Nos. WO 97/13845 and 98/27203, each of which is incorporated herein by reference. Particularly preferred host cells for purposes of the present invention are *Streptomyces* and *Saccharopolyspora* host cells, as discussed in greater detail below.

In a preferred embodiment, the expression vectors of the invention are  
20 used to construct a heterologous recombinant *Streptomyces* host cell that expresses a recombinant PKS of the invention. *Streptomyces* is a convenient host for expressing polyketides, because polyketides are naturally produced in certain *Streptomyces* species, and *Streptomyces* cells generally produce the precursors needed to form the desired polyketide. Those of skill in the art will recognize that,  
25 if a *Streptomyces* host cell produces any portion of a PKS enzyme or produces a polyketide modification enzyme, the recombinant vector need drive expression of only those genes constituting the remainder of the desired PKS enzyme or other polyketide-modifying enzymes. Thus, such a vector may comprise only a single ORF, with the desired remainder of the polypeptides constituting the PKS  
30 provided by the genes on the host cell chromosomal DNA.

If a *Streptomyces* or other host cell ordinarily produces polyketides, it may be desirable to modify the host so as to prevent the production of endogenous polyketides prior to its use to express a recombinant PKS of the invention. Such

modified hosts include *S. coelicolor* CH999 and similarly modified *S. lividans* described in U.S. Patent No. 5,672,491, and PCT publication Nos. WO 95/08548 and WO 96/40968, incorporated herein by reference. In such hosts, it may not be necessary to provide enzymatic activities for all of the desired post-translational  
5 modifications of the enzymes that make up the recombinantly produced PKS, because the host naturally expresses such enzymes. In particular, these hosts generally contain holo-ACP synthases that provide the phosphopantotheinyl residue needed for functionality of the PKS.

The invention provides a wide variety of expression vectors for use in  
10 *Streptomyces*. The replicating expression vectors of the present invention include, for example and without limitation, those that comprise an origin of replication from a low copy number vector, such as SCP2\* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and  
15 Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and pSG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, *Plasmid* 30: 131-140, each of which is  
20 incorporated herein by reference). For non-replicating and integrating vectors and generally for any vector, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p11, and pBR. For phage based vectors, the  
25 phage phiC31 and its derivative KC515 can be employed (see Hopwood *et al.*, *supra*). Also, plasmid pSET152, plasmid pSAM, plasmids pSE101 and pSE211, all of which integrate site-specifically in the chromosomal DNA of *S. lividans*, can be employed for purposes of the present invention.

The *Streptomyces* recombinant expression vectors of the invention  
30 typically comprise one or more selectable markers, including antibiotic resistance conferring genes selected from the group consisting of the *ermE* (confers resistance to erythromycin and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4*

(confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes. Alternatively, several polyketides are naturally colored, and this characteristic can provide a built-in marker for  
5 identifying cells.

Megalomicins are currently produced only by the relatively genetically intractable host *Micromonospora megalomicinea*. This bacteria has not been commonly used in the fermentation industry for the large-scale production of antibiotics, and methods for high level production of megalomicin and its analogs  
10 are needed. In contrast, the streptomycete bacteria have been widely used for almost 50 years and are excellent hosts for production of megalomicin and its analogs. *Streptomyces lividans* and *S. coelicolor* have been developed for the expression of heterologous PKS systems. These organisms can stably maintain cloned heterologous PKS genes, express them at high levels under controlled  
15 conditions, and modify the corresponding PKS proteins (e.g., phosphopantotheinylation) so that they are capable of production of the polyketide they encode. Furthermore, these hosts contain the necessary pathways to produce the substrates required for polyketide synthesis; e.g. propionyl-CoA and methylmalonyl-CoA. A wide variety of cloning and expression vectors are  
20 available for these hosts, as are methods for the introduction and stable maintenance of large segments of foreign DNA. Relative to *Micromonospora* spp., *S. lividans* and *S. coelicolor* grow well on a number of media and have been adapted for high level production of polyketides in fermentors. If production levels are low, a number of rational approaches are available to improve yield (see  
25 Hosted and Baltz, 1996, *Trends Biotechnol.* 14(7):245-50, incorporated herein by reference). Empirical methods to increase the titers of these macrolides, long since proven effective for numerous bacterial polyketides, can also be employed.

Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which have been  
30 modified so as not to produce the polyketide actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent Nos. 5,830,750 and 6,022,731 and U.S. patent application Serial No. 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. These vectors are

particularly preferred in that they contain promoters compatible with numerous and diverse *Streptomyces* spp. Particularly useful promoters for *Streptomyces* host cells include those from PKS gene clusters that result in the production of polyketides as secondary metabolites, including promoters from aromatic (Type II) PKS gene clusters. Examples of Type II PKS gene cluster promoters are *act* gene promoters and *tcm* gene promoters; an example of a Type I PKS gene cluster promoter are the promoters of the spiramycin PKS genes and DEBS genes. The present invention also provides the megalomicin biosynthetic gene promoters in recombinant form. These promoters can be used to drive expression of the megalomicin biosynthetic genes or any other coding sequence of interest in host cells in which the promoter functions, particularly *Micromonospora megalomicea* and generally any *Streptomyces* species.

As described above, particularly useful control sequences are those that alone or together with suitable regulatory systems activate expression during transition from growth to stationary phase in the vegetative mycelium. The promoter contained in the aforementioned plasmid pRM5, i.e., the *actII/actIII* promoter pair and the *actIII-ORF4* activator gene, is particularly preferred. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actIII-ORF4* gene described above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833; supra).

To provide a preferred host cell and vector for purposes of the invention, the megalomicin biosynthetic genes are placed on a recombinant expression vector and transferred to the non-macrolide producing hosts *Streptomyces lividans* K4-114 and *S. coelicolor* CH999. Transformation of *S. lividans* K4-114 or *S. coelicolor* CH999 with this expression vector results in a strain which produces



detectable amounts of megalomicin as determined by analysis of extracts by LC/MS. As noted above, the present invention also provides recombinant DNA compounds in which the encoded megalomicin module I KS domain is inactivated (the KS1° mutation). The introduction into *Streptomyces lividans* or *S. coelicolor* of a recombinant expression vector of the invention that encodes a megalomicin PKS with a KS1° domain produces a host cell useful for making polyketides by a process known as diketide feeding. The resulting host cells can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare megalomicin derivatives. Such cells of the invention are especially useful in the production of 13-substituted-6-deoxyerythronolide B compounds in recombinant host cells. Preferred compounds of the invention include those compounds in which the substituent at the 13-position is propyl, vinyl, propargyl, other lower alkyl, and substituted alkyl. In a preferred embodiment, the meg PKS is produced from a recombinant construct in which the *megAIII* gene has been altered to abolish the regions of identical coding sequence it otherwise shares with the *megAI* gene, or a hybrid PKS is employed in which the *megAIII* gene product has been replaced by the *oleAIII* gene product. Recombinant *oleAIII* genes are described in, for example, PCT patent publication No. 00/026349 and U.S. patent application Serial No. 09/428,517, filed 28 Oct. 1999, both of which are incorporated herein by reference.

The recombinant host cells of the invention can express all of the megalomicin biosynthetic genes or only a subset of the same. For example, if only the genes for the megalomicin PKS are expressed in a host cell that otherwise does not produce polyketide modifying enzymes that can act on the polyketide produced, then the host cell produces unmodified polyketides, called macrolide aglycones. Such macrolide aglycones can be hydroxylated and glycosylated by adding them to the fermentation of a strain such as, for example, *Streptomyces antibioticus* or *Saccharopolyspora erythraea*, that contains the requisite modification enzymes.

There are a wide variety of diverse organisms that can modify macrolide aglycones to provide compounds with, or that can be readily modified to have, useful activities. For example, as shown in Figure 5, *Saccharopolyspora erythraea* can convert 6-dEB to a variety of useful compounds. The erythronolide 6-dEB is

converted by the *eryF* gene product to erythronolide B, which is, in turn, glycosylated by the *eryBV* gene product to obtain 3-O-mycarosylerythronolide B, which contains L-mycarose at C-3. The *eryCIII* gene product then converts this compound to erythromycin D by glycosylation with D-desosamine at C-5.

5 Erythromycin D, therefore, differs from 6-dEB through glycosylation and by the addition of a hydroxyl group at C-6. Erythromycin D can be converted to erythromycin B in a reaction catalyzed by the *eryG* gene product by methylating the L-mycarose residue at C-3. Erythromycin D is converted to erythromycin C by the addition of a hydroxyl group at C-12 in a reaction catalyzed by the *eryK* gene

10 product. Erythromycin A is obtained from erythromycin C by methylation of the mycarose residue in a reaction catalyzed by the *eryG* gene product. The unmodified megalomicin compounds provided by the present invention, such as, for example, the 6-dEB or 6-dEB analogs, produced in *Streptomyces lividans*, can be provided to cultures of *S. erythraea* and converted to the corresponding

15 derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in the examples below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol.* 164(1): 425-433). Also, one can employ other mutant strains, such as

20 *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production.

Moreover, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described

25 above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be chemically altered after fermentation. Thus, *Streptomyces venezuelae*, which produces picromycin, contains enzymes that can transfer a desosaminyl group to the C-5 hydroxyl and a

30 hydroxyl group to the C-12 position. In addition, *S. venezuelae* contains a glucosylation activity that glucosylates the 2'-hydroxyl group of the desosamine sugar. This latter modification reduces antibiotic activity, but the glucosyl residue is removed by enzymatic action prior to release of the polyketide from the cell.

Another organism, *S. narbonensis*, contains the same modification enzymes as *S. venezuelae*, except the C-12 hydroxylase. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. narbonensis* and *S. venezuelae*.

Other organisms suitable for making compounds of the invention include *Micromonospora megalomicea* (discussed above), *Streptomyces antibioticus*, *S. fradiae*, and *S. thermotolerans*. *S. antibioticus* produces oleandomycin and contains enzymes that hydroxylate the C-6 and C-12 positions, glycosylate the C-3 hydroxyl with oleandrose and the C-5 hydroxyl with desosamine, and form an epoxide at C-8-C-8a. *S. fradiae* contains enzymes that glycosylate the C-5 hydroxyl with mycaminose and then the 4'-hydroxyl of mycaminose with mycarose, forming a disaccharide. *S. thermotolerans* contains the same activities as *S. fradiae*, as well as acylation activities. Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the macrolide aglycones of the invention by action of the enzymes endogenous to *S. antibioticus*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *megAI*, *megAII*, and *megAIII* genes with one or more deletions and/or insertions, including replacements of a *megA* gene fragment with a gene fragment from a heterologous PKS gene (as discussed in the next Section), can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Streptomyces antibioticus*, *S. venezuelae*, *S. narbonensis*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans*.

A number of erythromycin high-producing strains of *Saccharopolyspora erythraea* and *Streptomyces fradiae* have been developed, and in a preferred embodiment, the megalomicin PKS and/or other megalomicin biosynthetic genes are introduced into such strains (or erythromycin non-producing mutants thereof) to provide the corresponding modified megalomicin compounds in high yields. Those of skill in the art will appreciate that *S. erythraea* contains the desosamine

and mycarose biosynthetic and transfer genes as well as DEBS, which, as noted above, makes the same macrolide aglycone, 6-dEB, as the megalomicin PKS. *S. erythraea* does not make megosamine or its corresponding transferase gene, and does not contain the acylation gene of *Micromonospora megalomicea*. Finally, the  
5 *S. erythraea eryG* gene product converts mycarose to cladinose, which does not occur in *M. megalomicea*. Thus, the present invention provides a wide variety of *S. erythraea* recombinant host cells, including, for example, those that contain:

- (i) wild-type erythromycin biosynthetic genes with recombinant megosamine biosynthetic and transfer genes, with and without megalomicin  
10 acylation genes;
- (ii) wild-type erythromycin biosynthetic genes except *eryG*, with recombinant megosamine biosynthetic and transfer genes, with and without megalomicin acylation genes; and
- (iii) as in (i) and (ii), except that the *eryA* genes are inactive or deleted and  
15 recombinant *megA* genes have been introduced.

The invention provides other *S. erythraea* strains as well, including those in which any one or more of the erythromycin biosynthetic genes have been deleted or otherwise rendered inactive and in which at least one megalomicin biosynthetic gene has been introduced.

20 For example, the present invention enables one to express the megosamine genes in a *Saccharopolyspora erythraea eryG* mutant in which the erythromycin C made by this mutant is converted to megalomicin A. Alternatively, one could use an erythromycin C high-producing strain of *S. erythraea* in biotransformation methods in which the erythromycin C is fed to a *Streptomyces lividans* strain  
25 carrying only the megosamine biosynthesis and glycosyltransferase genes. As another alternative, one could use a strain of *S. lividans* that carries suitable erythromycin production genes along with the daunosamine biosynthesis genes plus *geneX* and *geneY* of Figure 5, or all of the megosamine biosynthesis genes, to produce megalomicin A.

30 All or some of the megalomicin gene cluster can be easily cloned under control of a suitable promoter in pCK7 or pSET152 either in one or two plasmids and introduced into the *Saccharopolyspora erythraea eryG* mutant. The *actII-ORF4/actIp* system and the *phiC31/int* system in pSET function well in this

organism (see Rowe *et al.*, 1998, *Gene*, 216:215-23, incorporated herein by reference). Alternatively, the megosamine biosynthesis genes are introduced into *Streptomyces lividans* on the same plasmids and the production of megalomicin A or its precursor mediated by bioconversion, done by feeding erythronolide B, 3-  
5 alpha-mycarosylerythronolide B, erythromycin D or erythromycin C to the *S. lividans* strain.

Lack of adequate resistance to megalomicin A in *S. erythraea* or *S. lividans* is not expected, because both organisms have MLS resistance genes (*ermE* and *mgt/lrm*, respectively), which confer resistance to several 14-membered  
10 macrolides (see Cundliffe, 1989, *Annu. Rev. Microbiol.* 43:207-33; Jenkins and Cundliffe, 1991, *Gene* 108:55-62; and Cundliffe, 1992, *Gene*, 115:75-84, each of which is incorporated herein by reference). One can also readily determine the level of resistance of the *S. erythraea eryG* mutant and the *S. lividans* host cells to megalomicin A, both in plate tests and in liquid medium. One can repeat the  
15 bioconversion method using an *eryG* mutant of a high erythromycin A producing *S. erythraea* strain (or an *eryB* or *eryC* mutant, as necessary) to determine the level at which megalomicin A can be produced. Furthermore, if experience shows that high level megalomicin A production requires a higher level of resistance to this macrolide than present in *S. erythraea* or *S. lividans*, the necessary megalomicin  
20 self-resistance genes will be cloned from *M. megalomicea* and moved into either one of the heterologous hosts. This will be straightforward work since self-resistance genes are usually found in the cluster of macrolide biosynthesis genes and can be identified by their homology to known macrolide resistance genes and(or) by the resistance phenotype they impart to a strain that normally is  
25 sensitive.

Alternatively, *geneX* and *geneY* (Figure 5) can be added to cassettes containing the relevant daunosamine (*dnm*) biosynthesis genes (Figure 5) to provide the ability to make TDP-megosamine *in vivo* and attach it to an erythromycin alkycone. The TDP-daunosamine biosynthesis genes can be re-  
30 cloned from *Streptomyces peucetius* on two compatible and mutually selectable plasmids. When an *S. lividans* strain containing these two plasmids and the *dnmS* gene for TDP-daunosamine glycosyltransferase is grown in the presence of added epsilon-rhodomyconone, its glycoside with L-daunosamine, called rhodomycon D,

is produced in good yield. Thus, bioconversion of one of the erythromycins to megalomicin A should be observed when *geneX* and *geneY* are present. One can construct all five combination - the two *N*-dimethyltransferase genes and the three glycosyltransferase genes - to discriminate *geneX* and *geneY* from those connected  
5 with mycarose and desosamine biosynthesis and attachment in the megalomicin pathway.

Because the timing of megosamine addition is unknown, one can test erythronolide B, 3- $\alpha$ -mycarosylerythronolide B, erythromycin D and erythromycin C as substrates provided to a strain that expresses the megosamine  
10 biosynthetic and transferase genes. There is need to test the C3''' and(or) C4''' acylated metabolites like megalomicin C1, because these metabolites are made from megalomicin A and not the converse; based on the precedents in the biosynthesis of tylosin (see Arisawa *et al.*, 1994, *Appl. Environ. Microbiol.* 60: 2657-2661), carbomycin (see Epp *et al.*, 1989, *Gene* 85:293-301), and  
15 midecamycin (see Hara and Hutchinson, 1992, *J. Bacteriol.* 174, 5141-5144). If C-6 glycosylation of erythronolide B or 3- $\alpha$ -mycarosylerythronolide B (Figure 5) happens before addition of desosamine to C-5, then the erythromycin genes might not be able to complete formation of megalomicin A from some mono or diglycoside if the erythromycin glycosyltransferases cannot tolerate a C-6  
20 glycoside. Although unexpected, such an outcome could be circumvented in accordance with the methods of the invention by cloning further megalomicin biosynthesis genes into the appropriate *S. erythraea* background or into *S. lividans* – specifically, the necessary deoxysugar biosynthesis and attachment genes – to create a recombinant strain that produces megalomicin A.

25 The acyltransferase gene that adds acetate or propionate to the C3''' or C4''' positions of mycarose in megalomicin B, C1 and C2 (Figure 3) is contained within the cosmids of the invention and can be identified by scanning the sequence data for the megalomicin gene cluster to locate homologs of *carE* and *mdmB* or their *acyA* homologs from the tylosin producer. The *carE* and *acyA* genes govern  
30 C4''' acylation in the carbomycin and tylosin pathway, respectively. The megalomicin homolog has the equivalent function in megalomicin biosynthesis (but is specific for C3''' and C4''' acylation). The gene can be cloned under control of a suitable promoter and introduced into *S. lividans* to produce the



desired acyl derivative of megalomicin A. Alternatively, introduction of the *carE* gene can form megalomicin B. This gene can be cloned from the carbomycin, spiramycin or tylosin producers.

If the amount of megalomicin produced by an *S. erythraea* or *S. lividans* or  
5 other recombinant host cell is less than desired, yield can be improved by optimizing the growth medium and fermentation conditions, by increasing expression of the gene(s) that appear to be rate limiting, based on the level of pathway intermediates that are accumulated by the recombinant strain constructed, and by reconstructing the *ery*, *dnm*, and megalomicin biosynthesis genes on  
10 vectors like pSET152 that can be integrated into the genome to provide a stabler recombinant strain for strain improvement.

In another embodiment, the present invention provides recombinant vectors encoding one or more of the megosamine, desosamine, and mycarose biosynthetic and transfer genes and heterologous host cells comprising those  
15 vectors. In this embodiment of the invention, the heterologous host cell is typically a cell that is unable to produce the sugar and transfer it to a polyketide unless the vector of the invention is introduced. For example, neither *Streptomyces lividans* nor *S. coelicolor* is naturally capable of making megosamine, desosamine, or mycarose or transferring those moieties to a polyketide. However, the present  
20 invention provides recombinant *Streptomyces lividans* and *S. coelicolor* host cells that are capable of making megosamine, desosamine, and/or mycarose and transferring those moieties to a polyketide.

Moreover, additional recombinant gene products can be expressed in the host cell to improve production of a desired polyketide. As but one non-limiting  
25 example, certain of the recombinant PKS proteins of the invention may produce a polyketide other than or in addition to the predicted polyketide, because the polyketide is cleaved from the PKS by the thioesterase (TE) domain in module 6 prior to processing by other domains on the PKS, in particular, any KR, DH, and/or ER domains in module 6. The production of the predicted polyketide can  
30 be increased in such instances by deleting the TE domain coding sequences from the gene and, optionally, expressing the TE domain as a separate protein. See Gokhale *et al.*, Feb. 1999, "Mechanism and specificity of the terminal thioesterase

domain from the erythromycin polyketide synthase," *Chem. & Biol.* 6: 117-125, incorporated herein by reference.

Thus, in one important aspect, the present invention provides methods, expression vectors, and recombinant host cells that enable the production of megalomicin and hydroxylated and glycosylated derivatives of megalomicin in  
5 heterologous host cells. The present invention also provides methods for making a wide variety of polyketides derived in part from the megalomicin PKS or other biosynthetic genes, as described in the following Section.

#### 10 Section VI: Hybrid PKS Genes

The present invention provides recombinant DNA compounds encoding each of the domains of each of the modules of the megalomicin PKS as well as the other megalomicin biosynthetic enzymes. The availability of these compounds permits their use in recombinant procedures for production of desired portions of  
15 the megalomicin PKS fused to or expressed in conjunction with all or a portion of a heterologous PKS and, optionally, one or more polyketide modification enzymes. These compounds also permit the modification of polyketides with the various megalomicin modification enzymes. The resulting hybrid PKS can then be expressed in a host cell to produce a desired polyketide or modified form thereof.

20 Thus, in accordance with the methods of the invention, a portion of the megalomicin biosynthetic gene coding sequence that encodes a particular activity can be isolated and manipulated, for example, to replace the corresponding region in a different modular PKS gene or modification enzyme gene. In addition, coding sequences for individual proteins, modules, domains, and portions thereof of the  
25 megalomicin PKS can be ligated into suitable expression systems and used to produce the portion of the protein encoded. The resulting protein can be isolated and purified or can may be employed *in situ* to effect polyketide synthesis. Depending on the host for the recombinant production of the domain, module, protein, or combination of proteins, suitable control sequences such as promoters,  
30 termination sequences, enhancers, and the like are ligated to the nucleotide sequence encoding the desired protein in the construction of the expression vector, as described above.

In one important embodiment, the invention thus provides hybrid PKS enzymes and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the invention, a hybrid PKS is a recombinant PKS that comprises all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a first PKS and all or part of one or more extender modules, loading module, and/or thioesterase/cyclase domain of a second PKS. In one preferred embodiment, the first PKS is most but not all of the megalomicin PKS, and the second PKS is only a portion of a non-megalomicin PKS. An illustrative example of such a hybrid PKS includes a megalomicin PKS in which the megalomicin PKS loading module has been replaced with a loading module of another PKS. Another example of such a hybrid PKS is a megalomicin PKS in which the AT domain of extender module 3 is replaced with an AT domain that binds only malonyl CoA. In another preferred embodiment, the first PKS is most but not all of a non-megalomicin PKS, and the second PKS is only a portion of the megalomicin PKS. An illustrative example of such a hybrid PKS includes a rapamycin PKS in which an AT specific for malonyl CoA is replaced with the AT from the megalomicin PKS specific for methylmalonyl CoA. Other illustrative hybrid PKSs of the invention are described below.

Those of skill in the art will recognize that all or part of either the first or second PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See PCT patent application No. WO US99/15047, and Lau *et al.*, *infra*, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. Thus, the desired derivative coding sequences can be synthesized using standard solid phase synthesis methods such as those described by Jaye *et al.*, 1984, *J. Biol. Chem.* 259: 6331, and instruments for automated synthesis are available commercially from, for example, Applied Biosystems, Inc. For purposes of the invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

With this general background regarding hybrid PKSs of the invention, one can better appreciate the benefit provided by the DNA compounds of the invention

that encode the individual domains, modules, and proteins that comprise the megalomicin PKS. As described above, the megalomicin PKS is comprised of a loading module, six extender modules composed of a KS, AT, ACP, and zero, one, two, or three KR, DH, and ER domains, and a thioesterase domain. The DNA compounds of the invention that encode these domains individually or in combination are useful in the construction of the hybrid PKS encoding DNA compounds of the invention. For example, a DNA compound of the invention that encodes an extender module or portion of an extender module is useful in the construction of a coding sequence that encodes a protein subcomponent of a PKS.

10 The DNA compound of the invention that comprises a coding sequence of a PKS subunit protein is useful in the construction of an expression vector that drives expression of the subunit in a host cell that expresses the other subunits and so produces a functional PKS.

The recombinant DNA compounds of the invention that encode the loading module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS loading module is inserted into a DNA compound that comprises the coding sequence for one or more heterologous PKS extender modules. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by that for the coding sequence of the megalomicin PKS loading module provides a novel PKS. Examples include the DEBS, rapamycin, FK-506, FK-520, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS loading module is inserted into a DNA compound that comprises the coding sequence for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

15  
20  
25

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA (propionyl) specific AT with a malonyl CoA (acetyl), ethylmalonyl CoA (butyryl), or other CoA specific AT. In addition, the AT and/or ACP can be replaced by another AT and/or another ACP or an inactivated KS, such as a KS<sup>Q</sup>, an AT, and/or another

30

ACP. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

5 The recombinant DNA compounds of the invention that encode the first extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding  
10 sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the megalomicin PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the megalomicin PKS is  
15 inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a  
20 hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (which includes inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be  
25 replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a gene for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous first extender module coding sequence can  
30 be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

Those of skill in the art will recognize, however, that deletion of the KR domain of extender module 1 or insertion of a DH domain or DH and KR domains

into extender module 1 will prevent the typical cyclization of the polyketide at the hydroxyl group created by the KR if such hybrid module is employed as a first extender module in a hybrid PKS or is otherwise involved in producing a portion of the polyketide at which cyclization is to occur. Such deletions or insertions can  
5 be useful, however, to create linear molecules or to induce cyclization at another site in the molecule.

As noted above, the invention also provides recombinant PKSs and recombinant DNA compounds and vectors that encode such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are  
10 typically expressed in translational reading frame with the first two extender modules on a single protein, with the remaining modules and domains of a megalomicin, megalomicin derivative, or hybrid PKS expressed as one or more, typically two, proteins to form the multi-protein functional PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the  
15 PKS encoded thereby can be fed or supplied with N-acylcysteamine thioesters of precursor molecules to prepare megalomicin derivative compounds. See U.S. patent application Serial No. 09/492,733, filed 27 Jan. 2000, and PCT publication Nos. WO 00/44717, 99/03986 and 97/02358, each of which is incorporated herein by reference.

20 The recombinant DNA compounds of the invention that encode the second extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS second extender module is inserted into a DNA compound that comprises the  
25 coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that  
30 encodes the second extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.



In another embodiment, a portion or all of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; replacing the KR with a KR, a KR and a DH, or a KR, DH, and ER; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

The recombinant DNA compounds of the invention that encode the third extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the megalomicin PKS is inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the inactive KR; and/or

replacing the KR with an active KR, or a KR and DH, or a KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a gene for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

10           The recombinant DNA compounds of the invention that encode the fourth extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS fourth extender module is inserted into a DNA compound that comprises the coding  
15           sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes  
20           the fourth extender module of the megalomicin PKS is inserted into a DNA compound that comprises coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

          In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a  
25           hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating any one, two, or all three of the ER, DH, and KR; and/or replacing any one, two, or all three of the ER, DH, and KR with either a KR, a DH and KR, or a KR, DH, and ER. In addition,  
30           the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS (except for the DH and ER domains), from a coding sequence

for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

5           The recombinant DNA compounds of the invention that encode the fifth extender module of the megalomicin PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS fifth  
10           extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes  
15           the fifth extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequence for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

          In another embodiment, a portion or all of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to  
20           create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting (or inactivating) the KR; inserting a DH or a DH and ER; and/or replacing the KR with another KR, a DH and KR, or a DH, KR, and ER. In addition, the KS and/or ACP can be replaced  
25           with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous fifth extender module coding  
30           sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

          The recombinant DNA compounds of the invention that encode the sixth extender module of the megalomicin PKS and the corresponding polypeptides

encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the megalomicin PKS sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the megalomicin PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the sixth extender module of the megalomicin PKS is inserted into a DNA compound that comprises the coding sequences for the megalomicin PKS or a recombinant megalomicin PKS that produces a megalomicin derivative.

In another embodiment, a portion or all of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting or inactivating the KR or replacing the KR with another KR, a KR and DH; or a KR, DH, and an ER; and/or inserting a DH or a DH and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the megalomicin PKS, from a coding sequence for a PKS that produces a polyketide other than megalomicin, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes megalomicin, a megalomicin derivative, or another polyketide.

The sixth extender module of the megalomicin PKS is followed by a thioesterase domain. This domain is important in the cyclization of the polyketide and its cleavage from the PKS. The present invention provides recombinant DNA compounds that encode hybrid PKS enzymes in which the megalomicin PKS is fused to a heterologous thioesterase or a heterologous PKS is fused to the megalomicin PKS thioesterase. Thus, for example, a thioesterase domain coding sequence from another PKS gene can be inserted at the end of the sixth (or other final) extender module coding sequence in recombinant DNA compounds of the

invention or the megalomicin PKS thioesterase can be similarly fused to a heterologous PKS. Recombinant DNA compounds encoding this thioesterase domain are useful in constructing DNA compounds that encode the megalomicin PKS, a PKS that produces a megalomicin derivative, and a PKS that produces a polyketide other than megalomicin or a megalomicin derivative.

Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:

(i) from fusions of heterologous domain (where heterologous means the domains in a module are derived from at least two different naturally occurring modules) coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS, but also:

(ii) from fusions of heterologous modules (where heterologous module means two modules are adjacent to one another that are not adjacent to one another in naturally occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,

(iii) from expression of one or more megalomicin PKS genes with one or more non-megalomicin PKS genes, including both naturally occurring and recombinant non-megalomicin PKS genes, and

(iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described herein.

An example of a hybrid PKS comprising fused modules results from fusion of the loading module of either the DEBS PKS or the narbonolide PKS (see PCT patent application No. US99/11814, incorporated herein by reference) with extender modules 1 and 2 of the megalomicin PKS to produce a hybrid *megAI* gene. Co-expression of either one of these two hybrid *megAI* genes with the *megAII* and *megAIII* genes in suitable host cells, such as *Streptomyces lividans*, results in expression of a hybrid PKS of the invention that produces 6-deoxyerythronolide B (the polyketide product of the natural *megA* genes) in recombinant host cells. Co-expression of either one of these two hybrid *megAI*

genes with the *eryAII* and *eryAIII* genes similarly results in the production of 6-dEB, while co-expression with the analogous narbonolide PKS genes, *picAII*, *picAIII* and *picAIV*, results in the production of 3-deoxy-3-oxo-6-dEB (3-keto-6-dEB), useful in the production of ketolides, compounds with potent anti-bacterial activity.

Another example of a hybrid PKS comprising a hybrid module is prepared by co-expressing the *megAI* and *megAII* genes with a *megAIII* hybrid gene encoding extender module 5 and the KS and AT of extender module 6 of the megalomicin PKS fused to the ACP of module 6 and the TE of the narbonolide PKS. The resulting hybrid PKS of the invention produces 3-keto-6-dEB. This compound can also be prepared by a recombinant megalomicin derivative PKS of the invention in which the KR domain of module 6 of the megalomicin PKS has been deleted. Moreover, the invention provides hybrid PKSs in which not only the above changes have been made but also the AT domain of module 6 has been replaced with a malonyl-specific AT. These hybrid PKSs produce 2-desmethyl-3-deoxy-3-oxo-6-dEB, a useful intermediate in the preparation of 2-desmethyl ketolides, compounds with potent antibiotic activity.

Another illustrative example of a hybrid PKS includes the hybrid PKS of the invention resulting only from the latter change in the hybrid PKS just described. Thus, co-expression of the *megAI* and *megAII* genes with a hybrid *megAIII* gene in which the AT domain of module 6 has been replaced by a malonyl-specific AT results in the expression of a hybrid PKS that produces 2-desmethyl-6-dEB in recombinant host cells. This compound is a useful intermediate for making 2-desmethyl erythromycins in recombinant host cells of the invention, as well as for making 2-desmethyl semi-synthetic ketolides.

While many of the hybrid PKSs described above are composed primarily of megalomicin PKS proteins, those of skill in the art recognize that the present invention provides many different hybrid PKSs, including those composed of only a small portion of the megalomicin PKS. For example, the present invention provides a hybrid PKS in which a hybrid *eryAI* gene that encodes the megalomicin PKS loading module fused to extender modules 1 and 2 of DEBS is coexpressed with the *eryAII* and *eryAIII* genes. The resulting hybrid PKS produces 6-dEB, the product of the native DEBS. When the construct is expressed in



*Saccharopolyspora erythraea* host cells (either via chromosomal integration in the chromosome or via a vector that encodes the hybrid PKS), the resulting recombinant host cell of the invention produces erythromycins. Another illustrative example is the hybrid PKS of the invention composed of the *megAI* and *eryAII* and *eryAIII* gene products. This construct is also useful in expressing erythromycins in *Saccharopolyspora erythraea* host cells. In a preferred embodiment, the *S. erythraea* host cells are *eryAI* mutants that do not produce 6-deoxyerythronolide B.

Another example is the hybrid PKS of the invention composed of the products of the *picAI* and *picAII* genes (the two proteins that comprise the loading module and extender modules 1 - 4, inclusive, of the narbonolide PKS) and the *megAIII* gene. The resulting hybrid PKS produces the macrolide aglycone 3-hydroxy-narbonolide in *Streptomyces lividans* host cells and the corresponding erythromycins in *Saccharopolyspora erythraea* host cells.

Each of the foregoing hybrid PKS enzymes of the invention, and the hybrid PKS enzymes of the invention generally, can be expressed in a host cell that also expresses a functional *oleP* gene product. The *oleP* gene encodes an oleandomycin modification enzyme, and expression of the gene together with a hybrid PKS of the invention provides the compounds of the invention in which a C-8 hydroxyl, a C-8a or C-8-C-8a epoxide is present.

Recombinant methods for manipulating modular PKS genes to make hybrid PKS enzymes are described in U.S. Patent Nos. 5,672,491; 5,843,718; 5,830,750; and 5,712,146; and in PCT publication Nos. 98/49315 and 97/02358, each of which is incorporated herein by reference. A number of genetic engineering strategies have been used with DEBS to demonstrate that the structures of polyketides can be manipulated to produce novel natural products, primarily analogs of the erythromycins (see the patent publications referenced *supra* and Hutchinson, 1998, *Curr Opin Microbiol.* 1:319-329, and Baltz, 1998, *Trends Microbiol.* 6:76-83, incorporated herein by reference). Because of the similar activity of the megalomicin PKS and DEBS (both PKS enzymes produce the macrolide aglycone 6-dEB), these methods can be readily applied to the recombinant megalomicin PKS genes of the invention.

These techniques include: (i) deletion or insertion of modules to control chain length, (ii) inactivation of reduction/dehydration domains to bypass beta-carbon processing steps, (iii) substitution of AT domains to alter starter and extender units, (iv) addition of reduction/dehydration domains to introduce catalytic activities, and (v) substitution of ketoreductase KR domains to control hydroxyl stereochemistry. In addition, engineered blocked mutants of DEBS have been used for precursor directed biosynthesis of analogs that incorporate synthetically derived starter units. For example, more than 100 novel polyketides were produced by engineering single and combinatorial changes in multiple modules of DEBS. Hybrid PKS enzymes based on DEBS with up to three catalytic domain substitutions were constructed by cassette mutagenesis, in which various DEBS domains were replaced with domains from the rapamycin PKS (see Schweke *et al.*, 1995, *Proc. Nat. Acad. Sci. USA* 92, 7839-7843, incorporated herein by reference) or one more of the DEBS KR domains was deleted. Functional single domain replacements or deletions were combined to generate DEBS enzymes with double and triple catalytic domain substitutions (see McDaniel *et al.*, 1999, *Proc. Nat. Acad. Sci. USA* 96, 1846-1851, incorporated herein by reference). By providing the analogous megalomicin/rapamycin hybrid PKS enzymes, the present invention provides alternative means to make these polyketides.

Methods for generating libraries of polyketides have been greatly improved by cloning PKS genes as a set of three or more mutually selectable plasmids, each carrying a different wild-type or mutant PKS gene, then introducing all possible combinations of the plasmids with wild-type, mutant, and hybrid PKS coding sequences into the same host (see U.S. patent application Serial No. 60/129,731, filed 16 Apr. 1999, and PCT Pub. No. 98/27203, each of which is incorporated herein by reference). This method can also incorporate the use of a KS1° mutant, which by mutational biosynthesis can produce polyketides made from diketide starter units (see Jacobsen *et al.*, 1997, *Science* 277, 367-369, incorporated herein by reference), as well as the use of a truncated gene that leads to 12-membered macrolides or an elongated gene that leads to 16-membered ketolides. Moreover, by utilizing in addition one or more vectors that encode glycosyl biosynthesis and transfer genes, such as those of the present invention for megosamine,

desosamine, oleandrose, cladinose, and/or mycarose (in any combination), a large collection of glycosylated polyketides can be prepared.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant hybrid PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the invention.

#### **Avermectin**

10 U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemadectin.

15 MacNeil *et al.*, 1992, *Gene 115*: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

#### **Candicidin (FR008)**

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

#### **Epothilone**

20 PCT Pub. No. 00/031247 to Kosan.

#### **Erythromycin**

PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

25 Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.

#### Glycosylation Enzymes

PCT Pub. No. 97/23630 to Abbott.

#### **30 FK-506**

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506, *Eur. J. biochem.* 256: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506, *Eur. J. Biochem.* 244: 74-80.

Methyltransferase

- 5 US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520, *J. Bacteriol.* 178: 5243-5248.

10 **FK-520**

PCT Pub. No. 00/20601 to Kosan.

See also Nielsen *et al.*, 1991, *Biochem.* 30:5789-96 (enzymology of pipecolate incorporation).

**Lovastatin**

- 15 U.S. Pat. No. 5,744,350 to Merck.

**Narbomycin (and Picromycin)**

PCT Pub. No. WO US99/61599 to Kosan.

**Nemadectin**

MacNeil *et al.*, 1993, *supra*.

20 **Niddamycin**

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol.* 179: 7515-7522.

**Oleandomycin**

- 25 Swan *et al.*, 1994, Characterization of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

PCT Pub. No. 00/026349 to Kosan.

- 30 Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

**Platenolide**

EP Pub. No. 791,656 to Lilly.

### Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

- 5       Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

### Rifamycin

- 10       August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amiclatopsis mediterranei* S669, *Chemistry & Biology*, 5(2): 69-79.

### Soraphen

U.S. Pat. No. 5,716,849 to Novartis.

- 15       Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

### Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

- 20       Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

### Tylosin

EP Pub. No. 791,655 to Lilly.

- 25       Kuhstoss *et al.*, 1996, *Gene* 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

U.S. Pat. No. 5,876,991 to Lilly.

### Tailoring enzymes

Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355.

- 30       Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of PKS genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid PKS-encoding DNA compounds of the invention.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a cognate KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units" *Biochemistry* 38(5):1643-1651, incorporated herein by reference. One can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases", *Science* 284: 482-485, incorporated herein by reference.

The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Even where only two genes are used, there are often two or more modules in the hybrid gene in which all or part of the module is derived from a second (or third) PKS gene. Thus, as one illustrative example, the invention provides a hybrid PKS that contains the naturally occurring loading module and thioesterase domain as well as extender modules one, two, four, and six of the megalomicin PKS and further contains hybrid or heterologous extender modules three and five. Hybrid or heterologous extender modules three and five contain AT domains specific for malonyl CoA and derived from, for example, the rapamycin PKS genes.

The invention also provides libraries of PKS genes, PKS proteins, and ultimately, of polyketides, that are constructed by generating modifications in the megalomicin PKS so that the protein complexes produced have altered activities in one or more respects and thus produce polyketides other than the natural product of the PKS. Novel polyketides may thus be prepared, or polyketides in general prepared more readily, using this method. By providing a large number of



different genes or gene clusters derived from a naturally occurring PKS gene cluster, each of which has been modified in a different way from the native cluster, an effectively combinatorial library of polyketides can be produced as a result of the multiple variations in these activities. As will be further described below, the  
5 metes and bounds of this embodiment of the invention can be described on the polyketide, protein, and the encoding nucleotide sequence levels.

As described above, a modular PKS "derived from" the megalomicin or other naturally occurring PKS includes a modular PKS (or its corresponding encoding gene(s)) that retains the scaffolding of the utilized portion of the  
10 naturally occurring gene. Not all modules need be included in the constructs; however, the constructs can also comprise more than six modules. On the constant scaffold, at least one enzymatic activity is mutated, deleted, replaced, or inserted so as to alter the activity of the resulting PKS relative to the original (native) PKS. Alteration results when these activities are deleted or are replaced by a different  
15 version of the activity, or simply mutated in such a way that a polyketide other than the natural product results from these collective activities. This occurs because there has been a resulting alteration of the starter unit and/or extender unit, stereochemistry, chain length or cyclization, and/or reductive or dehydration cycle outcome at a corresponding position in the product polyketide. Where a  
20 deleted activity is replaced, the origin of the replacement activity may come from a corresponding activity in a different naturally occurring PKS or from a different region of the megalomicin PKS. Any or all of the megalomicin PKS genes may be included in the derivative or portions of any of these may be included, but the scaffolding of a functional PKS protein is retained in whatever derivative is  
25 constructed. The derivative preferably contains a thioesterase activity from the megalomicin or another PKS.

Thus, a PKS derived from the megalomicin PKS includes a PKS that contains the scaffolding of all or a portion of the megalomicin PKS. The derived PKS also contains at least two extender modules that are functional, preferably  
30 three extender modules, and more preferably four or more extender modules, and most preferably six extender modules. The derived PKS also contains mutations, deletions, insertions, or replacements of one or more of the activities of the functional modules of the megalomicin PKS so that the nature of the resulting

polyketide is altered at both the protein and DNA sequence levels. Particular preferred embodiments include those wherein a KS, AT, or ACP domain has been deleted or replaced by a version of the activity from a different PKS or from another location within the same PKS. Also preferred are derivatives where at least one non-condensation cycle enzymatic activity (KR, DH, or ER) has been deleted or added or wherein any of these activities has been mutated so as to change the structure of the polyketide synthesized by the PKS.

Conversely, also included within the definition of a PKS derived from the megalomicin PKS are functional non-megalomicin PKS modules or their encoding genes wherein at least one domain or coding sequence therefor of a megalomicin PKS module has been inserted. Exemplary is the use of the megalomicin AT for extender module 2, which accepts a methylmalonyl CoA extender unit rather than malonyl CoA, to replace a malonyl specific AT in another PKS. Other examples include insertion of portions of non-condensation cycle enzymatic activities or other regions of megalomicin synthase activity into a heterologous PKS at both the DNA and protein levels.

Thus, there are at least five degrees of freedom for constructing a hybrid PKS in terms of the polyketide that will be produced. First, the polyketide chain length is determined by the number of extender modules in the PKS, and the present invention includes hybrid PKSs that contain 6, as well as fewer or more than 6, extender modules. Second, the nature of the carbon skeleton of the PKS is determined by the specificities of the acyl transferases that determine the nature of the extender units at each position, e.g., malonyl, methylmalonyl, ethylmalonyl, or other substituted malonyl. Third, the loading module specificity also has an effect on the resulting carbon skeleton of the polyketide. The loading module may use a different starter unit, such as acetyl, butyryl, and the like. As noted above, another method for varying loading module specificity involves inactivating the KS activity in extender module 1 (KS1) and providing alternative substrates, called diketides, that are chemically synthesized analogs of extender module 1 diketide products, for extender module 2. This approach was illustrated in PCT publication Nos. 97/02358 and 99/03986, incorporated herein by reference, wherein the KS1 activity was inactivated through mutation. Fourth, the oxidation state at various positions of the polyketide will be determined by the dehydratase and reductase

portions of the modules. This will determine the presence and location of ketone and alcohol moieties and C-C double bonds or C-C single bonds in the polyketide.

Finally, the stereochemistry of the resulting polyketide is a function of three aspects of the synthase. The first aspect is related to the AT/KS specificity associated with substituted malonyls as extender units, which affects  
5 stereochemistry only when the reductive cycle is missing or when it contains only a ketoreductase, as the dehydratase would abolish chirality. Second, the specificity of the ketoreductase may determine the chirality of any beta-OH. Finally, the enoylreductase specificity for substituted malonyls as extender units may influence  
10 the stereochemistry when there is a complete KR/DH/ER available.

Thus, the modular PKS systems generally and the megalomicin PKS system particularly permit a wide range of polyketides to be synthesized. As compared to the aromatic PKS systems, the modular PKS systems accept a wider range of starter units, including aliphatic monomers (acetyl, propionyl, butyryl,  
15 isovaleryl, and the like.), aromatics (aminohydroxybenzoyl), alicyclics (cyclohexanoyl), and heterocyclics (thiazolyl). Certain modular PKSs have relaxed specificity for their starter units (Kao *et al.*, 1994, *Science, supra*). Modular PKSs also exhibit considerable variety with regard to the choice of extender units in each condensation cycle. The degree of beta-ketoreduction following a  
20 condensation reaction can be altered by genetic manipulation (Donadio *et al.*, 1991, *Science, supra*; Donadio *et al.*, 1993, *Proc. Natl. Acad. Sci. USA* 90: 7119-7123). Likewise, the size of the polyketide product can be varied by designing mutants with the appropriate number of modules (Kao *et al.*, 1994, *J. Am. Chem. Soc.* 116:11612-11613). Lastly, modular PKS enzymes are particularly well  
25 known for generating an impressive range of asymmetric centers in their products in a highly controlled manner. The polyketides, antibiotics, and other compounds produced by the methods of the invention are typically single stereoisomeric forms. Although the compounds of the invention can occur as mixtures of stereoisomers, it may be beneficial in some instances to generate individual  
30 stereoisomers. Thus, the combinatorial potential within modular PKS pathways based on any naturally occurring modular, such as the megalomicin, PKS scaffold is virtually unlimited.

While hybrid PKSs are most often produced by "mixing and matching" portions of PKS coding sequences, mutations in DNA encoding a PKS can also be used to introduce, alter, or delete an activity in the encoded polypeptide. Mutations can be made to the native sequences using conventional techniques. The substrates for mutation can be an entire cluster of genes or only one or two of them; the substrate for mutation may also be portions of one or more of these genes. Techniques for mutation include preparing synthetic oligonucleotides including the mutations and inserting the mutated sequence into the gene encoding a PKS subunit using restriction endonuclease digestion. See, e.g., Kunkel, 1985, *Proc. Natl. Acad. Sci. USA* 82: 448; Geisselsoder *et al.*, 1987, *BioTechniques* 5:786. Alternatively, the mutations can be effected using a mismatched primer (generally 10-20 nucleotides in length) that hybridizes to the native nucleotide sequence, at a temperature below the melting temperature of the mismatched duplex. The primer can be made specific by keeping primer length and base composition within relatively narrow limits and by keeping the mutant base centrally located. See Zoller and Smith, 1983, *Methods Enzymol.* 100:468. Primer extension is effected using DNA polymerase, the product cloned, and clones containing the mutated DNA, derived by segregation of the primer extended strand, selected. Identification can be accomplished using the mutant primer as a hybridization probe. The technique is also applicable for generating multiple point mutations. See, e.g., Dalbie-McFarland *et al.*, 1982, *Proc. Natl. Acad. Sci. USA* 79: 6409. PCR mutagenesis can also be used to effect the desired mutations.

Random mutagenesis of selected portions of the nucleotide sequences encoding enzymatic activities can also be accomplished by several different techniques known in the art, e.g., by inserting an oligonucleotide linker randomly into a plasmid, by irradiation with X-rays or ultraviolet light, by incorporating incorrect nucleotides during *in vitro* DNA synthesis, by error-prone PCR mutagenesis, by preparing synthetic mutants, or by damaging plasmid DNA *in vitro* with chemicals, in accordance with the methods of the present invention. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, nitrosoguanidine, hydroxylamine, agents which damage or remove bases thereby preventing normal base-pairing such as hydrazine or formic acid, analogues of nucleotide precursors such as 5-bromouracil, 2-aminopurine, or acridine

intercalating agents such as proflavine, acriflavine, quinacrine, and the like.

Generally, plasmid DNA or DNA fragments are treated with chemical mutagens, transformed into *E. coli* and propagated as a pool or library of mutant plasmids.

In constructing a hybrid PKS of the invention, regions encoding enzymatic activity, i.e., regions encoding corresponding activities from different PKS  
5 synthases or from different locations in the same PKS, can be recovered, for example, using PCR techniques with appropriate primers. By "corresponding" activity encoding regions is meant those regions encoding the same general type of activity. For example, a KR activity encoded at one location of a gene cluster  
10 "corresponds" to a KR encoding activity in another location in the gene cluster or in a different gene cluster. Similarly, a complete reductase cycle could be considered corresponding. For example, KR/DH/ER can correspond to a KR alone.

If replacement of a particular target region in a host PKS is to be made,  
15 this replacement can be conducted *in vitro* using suitable restriction enzymes. The replacement can also be effected *in vivo* using recombinant techniques involving homologous sequences framing the replacement gene in a donor plasmid and a receptor region in a recipient plasmid. Such systems, advantageously involving plasmids of differing temperature sensitivities are described, for example, in PCT  
20 publication No. WO 96/40968, incorporated herein by reference. The vectors used to perform the various operations to replace the enzymatic activity in the host PKS genes or to support mutations in these regions of the host PKS genes can be chosen to contain control sequences operably linked to the resulting coding sequences in a manner such that expression of the coding sequences can be  
25 effected in an appropriate host.

However, simple cloning vectors may be used as well. If the cloning vectors employed to obtain PKS genes encoding derived PKS lack control sequences for expression operably linked to the encoding nucleotide sequences, the nucleotide sequences are inserted into appropriate expression vectors. This  
30 need not be done individually, but a pool of isolated encoding nucleotide sequences can be inserted into expression vectors, the resulting vectors transformed or transfected into host cells, and the resulting cells plated out into individual colonies. The invention provides a variety of recombinant DNA

compounds in which the various coding sequences for the domains and modules of the megalomicin PKS are flanked by non-naturally occurring restriction enzyme recognition sites.

The various PKS nucleotide sequences can be cloned into one or more recombinant vectors as individual cassettes, with separate control elements, or under the control of, e.g., a single promoter. The PKS subunit encoding regions can include flanking restriction sites to allow for the easy deletion and insertion of other PKS subunit encoding sequences so that hybrid PKSs can be generated. The design of such unique restriction sites is known to those of skill in the art and can be accomplished using the techniques described above, such as site-directed mutagenesis and PCR.

The expression vectors containing nucleotide sequences encoding a variety of PKS enzymes for the production of different polyketides are then transformed into the appropriate host cells to construct the library. In one straightforward approach, a mixture of such vectors is transformed into the selected host cells and the resulting cells plated into individual colonies and selected to identify successful transformants. Each individual colony has the ability to produce a particular PKS synthase and ultimately a particular polyketide. Typically, there will be duplications in some, most, or all of the colonies; the subset of the transformed colonies that contains a different PKS in each member colony can be considered the library. Alternatively, the expression vectors can be used individually to transform hosts, which transformed hosts are then assembled into a library. A variety of strategies are available to obtain a multiplicity of colonies each containing a PKS gene cluster derived from the naturally occurring host gene cluster so that each colony in the library produces a different PKS and ultimately a different polyketide. The number of different polyketides that are produced by the library is typically at least four, more typically at least ten, and preferably at least 20, and more preferably at least 50, reflecting similar numbers of different altered PKS gene clusters and PKS gene products. The number of members in the library is arbitrarily chosen; however, the degrees of freedom outlined above with respect to the variation of starter, extender units, stereochemistry, oxidation state, and chain length enables the production of quite large libraries.



Methods for introducing the recombinant vectors of the invention into suitable hosts are known to those of skill in the art and typically include the use of  $\text{CaCl}_2$  or agents such as other divalent cations, lipofection, DMSO, protoplast transformation, conjugation, infection, transfection, and electroporation. The polyketide producing colonies can be identified and isolated using known techniques and the produced polyketides further characterized. The polyketides produced by these colonies can be used collectively in a panel to represent a library or may be assessed individually for activity.

The libraries of the invention can thus be considered at four levels: (1) a multiplicity of colonies each with a different PKS encoding sequence; (2) the proteins produced from the coding sequences; (3) the polyketides produced from the proteins assembled into a functional PKS; and (4) antibiotics or compounds with other desired activities derived from the polyketides. Of course, combination libraries can also be constructed wherein members of a library derived, for example, from the megalomicin PKS can be considered as a part of the same library as those derived from, for example, the rapamycin PKS or DEBS.

Colonies in the library are induced to produce the relevant synthases and thus to produce the relevant polyketides to obtain a library of polyketides. The polyketides secreted into the media can be screened for binding to desired targets, such as receptors, signaling proteins, and the like. The supernatants *per se* can be used for screening, or partial or complete purification of the polyketides can first be effected. Typically, such screening methods involve detecting the binding of each member of the library to receptor or other target ligand. Binding can be detected either directly or through a competition assay. Means to screen such libraries for binding are well known in the art and can be applied in accordance with the methods of the present invention. Alternatively, individual polyketide members of the library can be tested against a desired target. In this event, screens wherein the biological response of the target is measured can more readily be included. Antibiotic activity can be verified using typical screening assays such as those set forth in Lehrer *et al.*, 1991, *J. Immunol. Meth.* 137:167-173, incorporated herein by reference, and in the Examples below.

The invention provides methods for the preparation of a large number of polyketides. These polyketides are useful intermediates in formation of

compounds with antibiotic or other activity through hydroxylation, epoxidation, and glycosylation reactions as described above. In general, the polyketide products of the PKS must be further modified, typically by hydroxylation and glycosylation, to exhibit potent antibiotic activity. Hydroxylation results in the novel polyketides  
5 of the invention that contain hydroxyl groups at C-6, which can be accomplished using the hydroxylase encoded by the *eryF* gene, and/or C-12, which can be accomplished using the hydroxylase encoded by the *picK* or *eryK* gene. Also, the *oleP* gene is available in recombinant form, which can be used to express the *oleP* gene product in any host cell. A host cell, such as a *Streptomyces* host cell or a  
10 *Saccharopolyspora erythraea* host cell, modified to express the *oleP* gene thus can be used to produce polyketides comprising the C-8-C-8a epoxide present in oleandomycin. Thus the invention provides such modified polyketides. The presence of hydroxyl groups at these positions can enhance the antibiotic activity of the resulting compound relative to its unhydroxylated counterpart.

15 Methods for glycosylating polyketides are generally known in the art and can be applied in accordance with the methods of the present invention; the glycosylation may be effected intracellularly by providing the appropriate glycosylation enzymes or may be effected *in vitro* using chemical synthetic means as described herein and in PCT publication No. WO 98/49315, incorporated  
20 herein by reference. Preferably, glycosylation with desosamine, mycarose, and/or megosamine is effected in accordance with the methods of the invention in recombinant host cells provided by the invention. In general, the approaches to effecting glycosylation mirror those described above with respect to hydroxylation. The purified enzymes, isolated from native sources or  
25 recombinantly produced may be used *in vitro*. Alternatively and as noted, glycosylation may be effected intracellularly using endogenous or recombinantly produced intracellular glycosylases. In addition, synthetic chemical methods may be employed.

The antibiotic modular polyketides may contain any of a number of  
30 different sugars, although D-desosamine, or a close analog thereof, is most common. Erythromycin, picromycin, megalomicin, narbomycin, and methymycin contain desosamine. Erythromycin also contains L-cladinose (3-O-methyl mycarose). Tylosin contains mycaminose (4-hydroxy desosamine), mycarose and

6-deoxy-D-allose. 2-acetyl-1-bromodesosamine has been used as a donor to glycosylate polyketides by Masamune *et al.*, 1975, *J. Am. Chem. Soc.* 97: 3512-3513. Other, apparently more stable donors include glycosyl fluorides, thioglycosides, and trichloroacetimidates; see Woodward *et al.*, 1981, *J. Am. Chem. Soc.* 103: 3215; Martin *et al.*, 1997, *J. Am. Chem. Soc.* 119: 3193; Toshima *et al.*, 1995, *J. Am. Chem. Soc.* 117: 3717; Matsumoto *et al.*, 1988, *Tetrahedron Lett.* 29: 3575. Glycosylation can also be effected using the polyketide aglycones as starting materials and using *Saccharopolyspora erythraea* or *Streptomyces venezuelae* or other host cell to make the conversion, preferably using mutants unable to synthesize macrolides, as discussed in the preceding Section.

Thus, a wide variety of polyketides can be produced by the hybrid PKS enzymes of the invention. These polyketides are useful as antibiotics and as intermediates in the synthesis of other useful compounds, as described in the following section.

#### Section VII: Host Cells Containing Multiple Expression Vectors

A recombinant host cell of the invention may contain nucleic acid encoding a megalomicin PKS domain, module, or protein, or megalomicin modification enzyme at a single genetic locus, *e.g.*, on a single plasmid or at a single chromosomal locus, or at different genetic loci, *e.g.*, on separate plasmids and/or chromosomal loci. By "multiple" is meant two or more; by "vector" is meant a nucleic acid molecule which can be used to transform host systems and which contains an independent expression system containing a coding sequence under control of a promoter and optionally a selectable marker and any other suitable sequences regulating expression. Typical such vectors are plasmids, but other vectors such as phagemids, cosmids, viral vectors and the like can be used according to the nature of the host. Of course, one or more of the separate vectors may integrate into the chromosome of the host (selection may not be required for maintenance of integrated vectors).

In one embodiment, the invention provides a recombinant host cell, which comprises at least two separate autonomously replicating recombinant DNA expression vectors, each of said vectors comprises a recombinant DNA compound encoding a megalomicin PKS domain or a megalomicin modification enzyme

operably linked to a promoter. In another embodiment, the invention provides a recombinant host cell, which comprises at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant  
5 DNA compound encoding a megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter. Preferably, the autonomously replicating recombinant DNA expression vector and/or the modified chromosome further comprises distinct selectable markers.

The above multiple-vector (chromosome) expression systems can also be  
10 used for expressing heterogeneous polyketide biosynthetic enzymes, *e.g.*, for expressing *Micromonospora megalomicea* megalomicin PKS protein, module, or domain or a megalomicin modification enzyme with a PKS protein, module, or domain, or modification enzyme from other origins in the same host cells. By placing various activities on different expression vectors, a high degree of  
15 variation can be achieved in an efficient manner. A variety of hosts can be used; any suitable host cell that can maintain multiple vectors can readily be used. Preferred hosts include *Streptomyces*, yeast, *E. coli*, other actinomycetes, and plant cells, and mammalian or insect cells or other suitable recombinant hosts can also be used. Preferred among yeast strains are *Saccharomyces cerevisiae* and *Pichia*  
20 *pastoris*. Preferred actinomycetes include various strains of *Streptomyces*.

If one chooses to use a host cell that does not naturally produce a polyketide, then one may need to ensure that the recombinant host is modified to also contain a holo ACP synthase activity that effects pantetheinylation of the acyl carrier protein. See PCT Pub. No. WO 97/13845, incorporated herein by  
25 reference. One of the multiple vectors may be used for this purpose. This activation step is necessary for activation of the ACP. The expression system for the holo ACP synthase may be supplied on a vector separate from that carrying a PKS coding sequence or may be supplied on the same vector or may be integrated into the chromosome of the host, or may be supplied as an expression system for a  
30 fusion protein with all or a portion of a polyketide synthase (see U.S. Patent No. 6,033,883, incorporated herein by reference).

It should be noted that in some recombinant hosts, it may also be necessary to activate the polyketides produced through postsynthesis modifications when

polyketides having such modifications are desired. If this is the case for a particular host, the host will be modified, for example by transformation, to contain those enzymes necessary for effecting these modifications. Among such enzymes, for example, are glycosylation enzymes. The use of multiple vectors can  
5 facilitate the introduction of expression systems for such enzymes.

In a preferred embodiment, the multiple vector system is used to assemble rapidly and efficiently a combinatorial library of polyketides and the PKS/modification enzymes that produce them. In an illustrative embodiment, the multiple vector system comprises four different vectors, one comprising the *megAI*  
10 gene, one the *megAII* gene, one the *megAIII* gene, and one the modification enzyme(s) gene(s). Each of these vectors can be modified to make a set of vectors. For example, one set could contain all possible AT substitutions in the loading and first and second extender modules of the *megAI* gene product. Another set could contain expression systems for a variety of different modification enzymes. With  
15 these four vectors sets and by combining each member of each set with each member of the other three sets, a very large library of cells, vector sets, and polyketides can be rapidly and efficiently assembled.

The combinatorial potential of a modular PKS such as the megalomicin PKS (ignoring the additional potential of different modification enzyme systems) is minimally given by:  $AT_L \times (AT_E \times 4)_M$  where  $AT_L$  is the number of loading  
20 acyl transferases,  $AT_E$  is the number of extender acyl transferases, and  $M$  is the number of modules in the gene cluster. The number 4 is present in the formula because this represents the number of ways a keto group can be modified by either 1) no reaction; 2) KR activity alone; 3) KR+DH activity; or 4) KR+DH+ER  
25 activity. It has been shown that expression of only the first two modules of the erythromycin PKS resulted in the production of a predicted truncated triketide product (See Kao et al., *J. Am. Chem. Soc.*, 116:11612-11613 ((1994))). A novel 12-membered macrolide similar to methymycin aglycone was produced by  
expression of modules 1-5 of this PKS in *S. coelicolor* (See Kao et al., *J. Am.*  
30 *Chem. Soc.*, 117:9105-9106 (1995)). This work shows that PKS modules are functionally independent so that lactone ring size can be controlled by the number of modules present.

In addition to controlling the number of modules, the modules can be genetically modified, for example, by the deletion of a ketoreductase domain as described by Donadio et al., *Science*, 252:675-679 (1991); and Donadio et al., *Gene*, 115:97-103 (1992). In addition, the mutation of an enoyl reductase domain  
5 was reported by Donadio, et al., *Proc. Natl. Acad. Sci.*, 90:7119-7123 (1993). These modifications also resulted in modified PKS and thus modified polyketides.

As stated above, in the present invention, the coding sequences for catalytic activities derived from the megalomicin PKS systems found in nature can be used in their native forms or modified by standard mutagenesis techniques to  
10 delete or diminish activity or to introduce an activity into a module in which it was not originally present. For example, a KR activity can be introduced into a module normally lacking that function.

In one embodiment of the invention herein, a single host cell is modified to contain a multiplicity of vectors, each vector contributing a portion of the  
15 synthesis of a megalomicin PKS and modification enzyme (if any) system. Each of the multiple vectors for production of the megalomicin PKS system typically encodes at least two modules, and at least one of the vectors integrates into the chromosome of the host. Integration can be effected using suitable phage or integrating vectors or by homologous recombination. If homologous  
20 recombination is used, the integration event may also be designed to delete endogenous PKS genes residing in the chromosome, as described in the PCT application WO 95/08548. In these embodiments, too, a selectable marker such as hygromycin or thiostrepton resistance can be included in the vector that effects integration.

25 As mentioned above, additional enzymes that effect post-translational modifications to the enzyme systems in the megalomicin PKS may be introduced into the host through suitable recombinant expression systems. In addition, enzymes that activate the polyketides themselves, for example, through glycosylation may be added. It may also be desirable to modify the cell to produce  
30 more of a particular substrate utilized in polyketide biosynthesis. For example, it is generally believed that malonyl CoA levels in yeast are higher than methylmalonyl CoA; if yeast is chosen as a host, it may be desirable to increase



methylmalonyl CoA levels by the addition of one or more biosynthetic enzymes therefor.

The multiple-vector expression system can also be used to make polyketides produced by the addition of synthetic starter units to a PKS that contains an inactivated ketosynthase (KS) in the first module. As noted above, this modification permits the system to incorporate a suitable diketide thioester such as 3-hydroxy-2-methyl pantonoic acid-N-acetyl cysteamine thioester, or similar thioesters of diketide analogs, as described by Jacobsen et al., *Science*, 277:367-369 (1997). The construction of PKS modules containing inactivated ketosynthase regions can be conducted by methods known in the art, such as the method described in U.S. Patent No. 6,080,555 and PCT publication Nos. WO 99/03986 and 97/02358, each of which is incorporated herein by reference, in accordance with the methods of the present invention.

The multiple-vector expression system can be used to produce polyketides in hosts that normally do not produce them, such as *E. coli* and yeast. It also provides more efficient means to provide a variety of polyketide products by supplying the elements of the introduced PKS, whether in an *E. coli* or yeast host or in other more traditionally used hosts, such as *Streptomyces*. The invention also includes libraries of polyketides prepared using the methods of the invention.

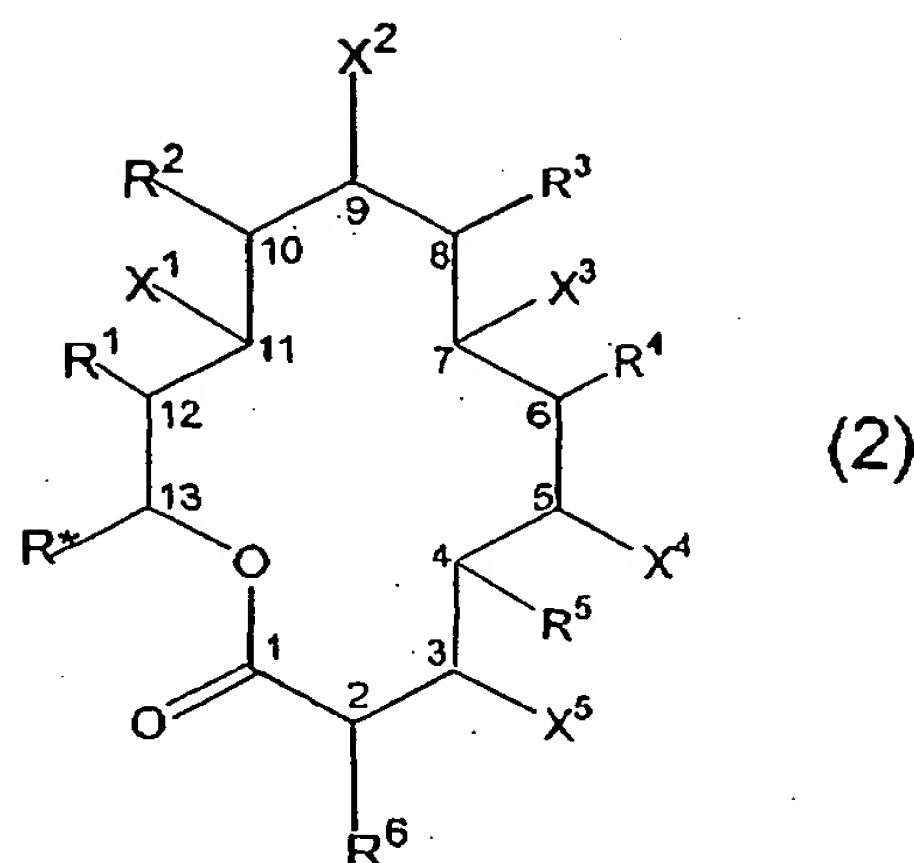
#### Section VIII: Compounds

The methods and recombinant DNA compounds of the invention are useful in the production of polyketides. In one important aspect, the invention provides methods for making antibiotic compounds related in structure to erythromycin, a potent antibiotic compound. The invention also provides novel ketolide compounds, polyketide compounds with potent antibiotic activity of significant interest due to activity against antibiotic resistant strains of bacteria. See Griesgraber et al., 1996, *J. Antibiot.* 49: 465-477, incorporated herein by reference. Most if not all of the ketolides prepared to date are synthesized using erythromycin A, a derivative of 6-dEB, as an intermediate. In one embodiment, the present invention provides the 3-keto derivatives of the megalomicins for use as antibiotics. In particular, the 3-keto derivative of megalomicin A is a preferred ketolide of the invention. These compounds can be made chemically, substantially

in accordance with the procedures for making ketolides described in the prior art, or in recombinant host cells of the invention in which the megosamine and desosamine biosynthetic and transferase genes are present but which do not make or transfer the mycarose moiety and/or the PKS has been modified to delete the KR domain of extender module 6. The invention also provides methods for making intermediates useful in preparing traditional, 6-dEB- and erythromycin-derived ketolide compounds. See Griesgraber *et al.*, *supra*; Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780; 5,444,051; 5,439,890; 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

As noted above, the hybrid PKS genes of the invention can be expressed in a host cell that contains the desosamine, megosamine, and/or mycarose biosynthetic genes and corresponding transferase genes as well as the required hydroxylase gene(s), which may, for example and without limitation, be either *picK*, *megK*, or *eryK* (for the C-12 position) and/or *megF* or *eryF* (for the C-6 position). The resulting compounds have antibiotic activity but can be further modified, as described in the patent publications referenced above, to yield a desired compound with improved or otherwise desired properties. Alternatively, the aglycone compounds can be produced in the recombinant host cell, and the desired glycosylation and hydroxylation steps carried out *in vitro* or *in vivo*, in the latter case by supplying the converting cell with the aglycone, as described above.

The compounds of the invention are thus optionally glycosylated forms of the polyketide set forth in formula (1) below which are hydroxylated at either the C-6 or the C-12 or both. The compounds of formula (1) can be prepared using the loading and the six extender modules of a modular PKS, modified or prepared in hybrid form as herein described. These polyketides have the formula:



including the glycosylated and isolated stereoisomeric forms thereof;

wherein R\* is a straight chain, branched or cyclic, saturated or unsaturated substituted or unsubstituted hydrocarbyl of 1-15C;

5 each of R<sup>1</sup>-R<sup>6</sup> is independently H or alkyl (1-4C) wherein any alkyl at R<sup>1</sup> may optionally be substituted;

each of X<sup>1</sup>-X<sup>5</sup> is independently two H, H and OH, or =O; or

each of X<sup>1</sup>-X<sup>5</sup> is independently H and the compound of formula (2)

contains a double-bond in the ring adjacent to the position of said X at 2-3, 4-5, 6-

10 7, 8-9 and/or 10-11;

with the proviso that:

at least two of R<sup>1</sup>-R<sup>6</sup> are alkyl (1-4C).

Preferred compounds comprising formula 2 are those wherein at least three of R<sup>1</sup>-R<sup>5</sup> are alkyl (1-4C), preferably methyl or ethyl; more preferably wherein at  
 15 least four of R<sup>1</sup>-R<sup>5</sup> are alkyl (1-4C), preferably methyl or ethyl. Also preferred are those wherein X<sup>2</sup> is two H, =O, or H and OH, and/or X<sup>3</sup> is H, and/or X<sup>1</sup> is OH and/or X<sup>4</sup> is OH and/or X<sup>5</sup> is OH. Also preferred are compounds with variable R\* when R<sup>1</sup>-R<sup>5</sup> is methyl, X<sup>2</sup> is =O, and X<sup>1</sup>, X<sup>4</sup> and X<sup>5</sup> are OH. The glycosylated forms (i.e., mycarose or cladinose at C-3, desosamine at C-5, and/or megosamine  
 20 at C-6) of the foregoing are also preferred.

As described above, there are a wide variety of diverse organisms that can modify compounds such as those described herein to provide compounds with or that can be readily modified to have useful activities. For example,

*Saccharopolyspora erythraea* can convert 6-dEB to a variety of useful

compounds. The compounds provided by the present invention can be provided to cultures of *Saccharopolyspora erythraea* and converted to the corresponding derivatives of erythromycins A, B, C, and D in accordance with the procedure provided in the Examples, below. To ensure that only the desired compound is produced, one can use an *S. erythraea eryA* mutant that is unable to produce 6-dEB but can still carry out the desired conversions (Weber *et al.*, 1985, *J. Bacteriol.* 164(1): 425-433). Also, one can employ other mutant strains, such as *eryB*, *eryC*, *eryG*, and/or *eryK* mutants, or mutant strains having mutations in multiple genes, to accumulate a preferred compound. The conversion can also be carried out in large fermentors for commercial production. Each of the erythromycins A, B, C, and D has antibiotic activity, although erythromycin A has the highest antibiotic activity. Moreover, each of these compounds can form, under treatment with mild acid, a C-6 to C-9 hemiketal with motilide activity. For formation of hemiketals with motilide activity, erythromycins B, C, and D, are preferred, as the presence of a C-12 hydroxyl allows the formation of an inactive compound that has a hemiketal formed between C-9 and C-12.

Thus, the present invention provides the compounds produced by hydroxylation and glycosylation of the compounds of the invention by action of the enzymes endogenous to *Saccharopolyspora erythraea* and mutant strains of *S. erythraea*. Such compounds are useful as antibiotics or as motilides directly or after chemical modification. For use as antibiotics, the compounds of the invention can be used directly without further chemical modification. Erythromycins A, B, C, and D all have antibiotic activity, and the corresponding compounds of the invention that result from the compounds being modified by *Saccharopolyspora erythraea* also have antibiotic activity. These compounds can be chemically modified, however, to provide other compounds of the invention with potent antibiotic activity. For example, alkylation of erythromycin at the C-6 hydroxyl can be used to produce potent antibiotics (clarithromycin is C-6-O-methyl), and other useful modifications are described in, for example, Griesgraber *et al.*, 1996, *J. Antibiot.* 49: 465-477, Agouridas *et al.*, 1998, *J. Med. Chem.* 41: 4080-4100, U.S. Patent Nos. 5,770,579; 5,760,233; 5,750,510; 5,747,467; 5,747,466; 5,656,607; 5,635,485; 5,614,614; 5,556,118; 5,543,400; 5,527,780;

5,444,051; 5,439,890; and 5,439,889; and PCT publication Nos. WO 98/09978 and 98/28316, each of which is incorporated herein by reference.

For use as motilides, the compounds of the invention can be used directly without further chemical modification. Erythromycin and certain erythromycin  
5 analogs are potent agonists of the motilin receptor that can be used clinically as prokinetic agents to induce phase III of migrating motor complexes, to increase esophageal peristalsis and LES pressure in patients with GERD, to accelerate gastric emptying in patients with gastric paresis, and to stimulate gall bladder contractions in patients after gallstone removal and in diabetics with autonomic  
10 neuropathy. See Peeters, 1999, Motilide Web Site, <http://www.med.kuleuven.ac.be/med/gih/motilid.htm>, and Omura *et al.*, 1987, Macrolides with gastrointestinal motor stimulating activity, *J. Med. Chem.* 30: 1941-3). The corresponding compounds of the invention that result from the compounds of the invention being modified by *Saccharopolyspora erythraea* also have motilide  
15 activity, particularly after conversion, which can also occur *in vivo*, to the C-6 to C-9 hemiketal by treatment with mild acid. Compounds lacking the C-12 hydroxyl are especially preferred for use as motilin agonists. These compounds can also be further chemically modified, however, to provide other compounds of the invention with potent motilide activity.

20 Moreover, and also as noted above, there are other useful organisms that can be employed to hydroxylate and/or glycosylate the compounds of the invention. As described above, the organisms can be mutants unable to produce the polyketide normally produced in that organism, the fermentation can be carried out on plates or in large fermentors, and the compounds produced can be  
25 chemically altered after fermentation. In addition to *Saccharopolyspora erythraea*, *Streptomyces venezuelae*, *S. narbonensis*, *S. antibioticus*, *Micromonospora megalomicea*, *S. fradiae*, and *S. thermotolerans* can also be used. In addition to antibiotic activity, compounds of the invention produced by treatment with *M. megalomicea* enzymes can have antiparasitic activity as well. Thus, the present  
30 invention provides the compounds produced by hydroxylation and glycosylation by action of the enzymes endogenous to *S. erythraea*, *S. venezuelae*, *S. narbonensis*, *S. antibioticus*, *M. megalomicea*, *S. fradiae*, and *S. thermotolerans*.

The present invention also provides methods and genetic constructs for producing the glycosylated and/or hydroxylated compounds of the invention directly in the host cell of interest. Thus, the recombinant genes of the invention, which include recombinant *megAI*, *megAII*, and *megAIII* genes with one or more  
5 deletions and/or insertions, including replacements of a *megA* gene fragment with a gene fragment from a heterologous PKS gene, can be included on expression vectors suitable for expression of the encoded gene products in *Saccharopolyspora erythraea*, *Micromonospora megalomicea*, *S. venezuelae*, *S. narbonensis*, *S. antibioticus*, *S. fradiae*, and *S. thermotolerans*.

10 The compounds of the invention can be produced by growing and fermenting the host cells of the invention under conditions known in the art for the production of other polyketides. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. The compounds can be readily formulated to provide the  
15 pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation will contain one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or  
20 parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use.

The carriers which can be used include water, glucose, lactose, gum acacia,  
25 gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl  
30 methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.



Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by  
5 reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases caused by infection, a compound of the invention may be administered orally, topically, parenterally, by  
10 inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

Dosage levels of the compounds of the invention are of the order from  
15 about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the invention may be administered on an intermittent  
20 basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral administration to humans may contain from 0.5 mg to 5 gm of  
25 active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention may be formulated within the range of, for example, 0.00001% to 60%  
30 by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight.

It will be understood, however, that the specific dose level for any particular patient will depend on a variety of factors. These factors include the

activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

- 5        A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the invention and shall not be construed as being a limitation on the scope of the invention or claims.

### Example 1

#### 10    Cloning and Characterization of the Megalomycin Biosynthetic Gene Cluster from *Micromonospora meglomicea*

##### **Experimental Procedures**

##### *Bacterial Strains, Media, and Growth Conditions*

- Routine DNA manipulations were performed in *Escherichia coli* XL1 Blue  
15    or *E. coli* XL1 Blue MR (Stratagene) using standard culture conditions (Sambrook  
*et al.*, 1989). *M. megalomicea* subs. *nigra* NRRL3275 was obtained from the  
ATCC collection and cultured according to recommended protocols. For isolation  
of genomic DNA, *M. megalomicea* was grown in TSB (Hopwood *et al.*, 1985) at  
30 °C. *S. lividans* K4-114 (Ziermann and Betlach, 1999), which carries a deletion  
20    of the actinorhodin biosynthetic gene cluster, was used as the host for expression  
of the *megAI-AIII* genes. *S. lividans* strains were maintained on R5 agar at 30°C  
and grown in liquid YEME for preparation of protoplasts (Hopwood *et al.*, 1985).  
*S. erythraea* NRRL2338 was used for expression of the megosamine genes. *S.*  
*erythraea* strains were maintained on R5 agar at 34°C and grown in liquid TSB for  
25    preparation of protoplasts.

##### *Manipulation of DNA and Organisms*

- Manipulation and transformation of DNA in *E. coli* was performed by  
standard procedures (Sambrook *et al.*, 1989) or by suppliers protocols. Protoplasts  
30    of *S. lividans* and *S. erythraea* were generated for transformation by plasmid DNA  
using the standard procedure. *S. lividans* transformants were selected on R5 using  
2 ml of a 0.5 mg/ml thiostrepton overlay. *S. erythraea* transformants were selected  
on R5 using 1.5 ml of a 0.6 mg/ml apramycin overlay.

*Isolation of the meg gene cluster*

A cosmid library was prepared in SuperCos (Stratagene) from *M. megalomicea* total DNA partially digested with *Sau3A* I, and introduced into *E. coli* using a Gigapack III XL (Stratagene) *in-vitro* packaging kit. <sup>32</sup>P-labelled DNA probes encompassing the KS2 domain from *ery* DEBS, or a mixture of segments encompassing modules 1 and 2 from *ery* DEBS were used separately to screen the cosmid library by colony hybridization. Several colonies which hybridized with the probes were further analyzed by sequencing the ends of their cosmid inserts using T3 and T7 primers. BLAST (Altschul *et al.*, 1990) analysis of the sequences revealed several colonies with DNA sequences highly homologous to genes from the *ery* cluster. Together with restriction analysis, this led to the isolation of two overlapping cosmids, pKOS079-93A and pKOS079-93D which covered ~45 kb of the *meg* cluster. A 400 bp PCR fragment was generated from the left end of and pKOS079-93D and used to reprobe the cosmid library. Likewise, a 200 bp PCR fragment generated from the right end of pKOS079-93A was used to reprobe the cosmid library. Analysis of hybridizing colonies as described above resulted in identification of two additional cosmids, pKOS079-138B and pKOS79-124B which overlap the previous two cosmids. BLAST analysis of the far left and right end sequences of these cosmids indicated no homology to any known genes related to polyketide biosynthesis and therefore indicates that the set of four cosmids spans the entire megalomicin biosynthetic gene cluster.

*DNA sequencing and analysis*

PCR-based double stranded DNA sequencing was performed on a Beckman CEQ 2000 capillary sequencer using reagents and protocols provided by the manufacturer. A shotgun library of the entire cosmid pKOS079-93D insert was made as follows: DNA was first digested with *Dra* I to eliminate the vector fragment, then partially digested with *Sau3A* I. After agarose electrophoresis, bands between 1-3 kb were excised from the gel and ligated with *Bam*H I digested pUC19. Another shotgun library was generated from a 12 kb *Xho* I/*Eco*R I fragment subcloned from cosmid pKOS079-93A to extend the sequence to the *megF* gene. A 4 kb *Bgl* II/*Xho* I fragment from cosmid pKOS079-138B was

sequenced by primer walking to extend the sequencing to the *megT* gene. Sequence was assembled using Sequencher (Gene Codes Corp.) software package and analyzed with MacVector (Oxford Molecular Group) and the NCBI BLAST server ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)).

5

### Plasmids

- Plasmid pKOS108-6 is a modified version of pKAO127'kan' (Ziermann and Betlach, 1999; Ziermann and Betlach, 2000) in which the *eryAI*-III genes between the *Pac* I and *EcoR* I sites have been replaced with the *megAI*-III genes.
- 10 This was done by first substituting a synthetic nucleotide DNA duplex (5'-TAAGAATTCGGAGATCTGGCCTCAGCTCTAGAC (SEQ ID NO: 21), complementary oligo 5'-AATTGTCTAGAGCTGAGGCCAGATCTCCGAATTCTTAAT (SEQ ID NO: 22)) between the *Pac* I and *EcoR* I sites of the pKAO127'kan' vector fragment.
- 15 The 22 kb *EcoR* I/*Bgl* II fragment from cosmid pKOS079-93D containing the *megAI*-II genes was inserted into *EcoR* I and *Bgl* II sites of the resulting plasmid to generate pKOS024-84. A 12 kb *Bgl* II/*BbvC* I fragment containing the *megAIII* and part of the *megCII* gene was subcloned from pKOS079-93A and excised as a *Bgl* II/*Xba* I fragment and ligated into the corresponding sites of pKOS024-84 to
- 20 yield the final expression plasmid pKOS108-06.

- The megosamine integrating vector, pKOS97-42, was constructed as follows: A subclone was generated containing the 4 kb *Xho* I/*Sca* I fragment from pKOS79-138B together with the 1.7 kb *Sca* I/*Pst* I fragment from pKOS79-93D in Litmus 28 (Stratagene). The entire 5.7 kb fragment was then excised as a *Spe* I/*Pst* I fragment and combined with the 6.3 kb *Pst* I/*EcoR* I fragment from KOS79-93D and *EcoR* I/*Xba* I digested pSET152 (Bierman *et al.*, 1992) to construct plasmid pKOS97-42.
- 25

### Production and analysis of secondary metabolites

- 30 Fermentation for production of polyketide, LC/MS analysis, and quantification of 6-dEB for *S. lividans* K4-114/pKOS108-6 and *S. lividans* K4-114/pKAO127'kan' were essentially as previously described (Xue *et al.*, 1999). *S. erythraea* NRRL2338 and *S. erythraea*/pKOS97-42 were grown for 6 days in F1

media (Brünker *et al.*, 1998). Samples of broth were clarified in a microcentrifuge (5 min, 13,000 rpm). For LC/MS preparation, isopropanol was added to the supernatant (1:2 ratio) and centrifuged again. Erythromycins and megalomicins were detected by electrospray mass spectrometry and quantity was determined by  
5 evaporative light scattering detection (ELSD). The LC retention time and mass spectra of erythromycin and megalomicins were identical to known standards.

#### *Nucleotide sequence of the meg gene cluster*

A series of 4 overlapping inserts containing the *meg* cluster (Figure 9) were  
10 isolated from a cosmid library prepared from total genomic DNA of *M. megalomicea* and covers > 100 kb of the genome. A contiguous 48 kb segment which encodes the megalomicin PKS and several deoxysugar biosynthetic genes was sequenced and analyzed. The segment contains 17 complete ORFs as well as an incomplete ORF at each end, organized as shown in Figure 9.

15 *PKS genes.* The ORFs *megAI*, *megAII* and *megAIII* encode the polyketide synthase responsible for synthesis of 6-dEB. The enzyme complex, *meg* DEBS, is highly similar to *ery* DEBS, with each of the three predicted polypeptides sharing an average of 83% overall similarity with their *ery* PKS counterpart. Both PKSs are composed of 6 modules (2 modules per polypeptide) and each module is  
20 organized in the identical manner (Figure 9). A dendrogram analysis (Schwecke *et al.*, 1995) employing 70 acyltransferase (AT) domains revealed that the 6 *meg* extender AT domains cluster with AT domains that incorporate methylmalonyl CoA (not shown). The loading module of *meg* DEBS also lacks a KS<sup>Q</sup> domain which is utilized by most macrolide PKSs for decarboxylation of the starter unit to  
25 initiate polyketide synthesis (Bisang *et al.*, 1999; Kuhstoss *et al.*, 1996; Kakavas *et al.*, 1997; Xue *et al.*, 1998), implying that priming begins with a propionate unit. In addition, a conserved Gly to Pro substitution in the NADPH-binding region of the ketoreductase (KR) domain of module 3 is observed in *meg* DEBS, which has been proposed to account for its inactivity in *ery* DEBS (Donadio *et al.*, 1991).

30 *Deoxysugar genes.* BLAST (Altschul *et al.*, 1990) analysis of the genes flanking the PKS indicated that 12 complete ORFs and 1 partial ORF appear to encode functions required for synthesis of one of the three megalomicin deoxysugars. Assignment of each ORF to a specific deoxysugar pathway was

made based on comparison to the *ery* genes and other related genes involved in deoxysugar biosynthesis (Table 2).

Table 2. Deduced functions of genes identified in the megalomicin gene cluster.

<i>Gene</i>	<i>Closest Match</i> ( <i>polypeptide</i> ) <sup>a</sup>	<i>% Sim</i> <sup>a</sup>	<i>Proposed</i> <i>Pathway</i>	<i>Proposed Function</i>	<i>Reference</i>
<i>megT</i>	EryBVI		Mycarose/ Megosamine	2,3-Dehydratase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megDVI</i>	EryCII	63	Megosamine	3,4-Isomerase	(Summers <i>et al.</i> , 1997)
<i>megDI</i>	EryCIII	79	Megosamine	Glycosyltransferase	(Summers <i>et al.</i> , 1997)
<i>megY</i>	AcyA ( <i>S.</i> <i>thermotolerans</i> )	52		Mycarose O-acyl- transferase	(Arisawa <i>et al.</i> , 1994)
<i>megDII</i>	EryCI	58	Megosamine	Aminotransferase	(Dhillon <i>et al.</i> , 1989; Summers <i>et al.</i> , 1997)
<i>megDIII</i>	DesVI ( <i>S.</i> <i>venezuelae</i> )	61	Megosamine	Dimethyltransferase	(Xue <i>et al.</i> , 1998)
<i>megDIV</i>	DmnU ( <i>S.</i> <i>peucetius</i> )	65	Megosamine	3,5-Epimerase	(Olano <i>et al.</i> , 1999)
<i>megDV</i>	Dehydrogenase ( <i>A. orientalis</i> )	61	Megosamine	4-Ketoreductase	(Summers <i>et al.</i> , 1997; van Wageningen <i>et al.</i> , 1998)
<i>megDVII</i>	EryBII	73	Megosamine	2,3-Reductase	(Summers <i>et al.</i> , 1997)
<i>megBV</i>	EryBV	86	Mycarose	Glycosyltransferase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megBIV</i>	EryBIV	80	Mycarose	4-Ketoreductase	(Summers <i>et al.</i> , 1997; Gaisser <i>et al.</i> , 1997)
<i>megAI</i>	EryAI	81	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megAII</i>	EryAII	85	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megAIII</i>	EryAIII	83	6-dEB	Polyketide Synthase	(Donadio and Katz, 1992)
<i>megCII</i>	EryCII	82	Desosamine	3,4-Isomerase	(Summers <i>et al.</i> , 1997)
<i>meg CIII</i>	EryCIII	89	Desosamine	Glycosyltransferase	(Summers <i>et al.</i> , 1997)
<i>megBII</i>	EryBII	87	Mycarose	2,3-Reductase	(Summers <i>et al.</i> , 1997)
<i>megH</i>	EryH	84		Thioesterase	(Haydock <i>et al.</i> , 1991)
<i>megF</i>	EryF			C-6 Hydroxylase	(Weber <i>et al.</i> , 1991)

5 a. Determined by BLASTX analysis using default parameters.



Three ORFs, *megBV*, *megCIII* and *megDI*, encode glycosyltransferases, apparently one for attachment of each deoxysugar to the macrolide. MegBV was most similar to EryBV, the erythromycin mycarosyltransferase, and hence was assigned to the mycarose pathway in the *meg* cluster. The closest match for both of the remaining glycosyltransferases was EryCIII, the desosaminytransferase in erythromycin biosynthesis. Given the higher degree of similarity between EryCIII and MegCIII (Table 2), MegCIII was designated the desosaminytransferase, leaving MegDI as the proposed megosaminytransferase. In similar fashion, assignments were made accordingly for; MegCII and MegDVI, two putative 3,4-isomerases similar to EryCII; MegBII and MegDVII, 2,3-reductases homologous to EryBII; MegBIV and MegDV, putative 4-ketoreductases similar to EryBIV (Table 2). The remaining ORFs involved in deoxysugar biosynthesis, *megT*, *megDII*, *megDIII* and *megDIV*, each encode a putative 2,3-dehydratase, aminotransferase, dimethyltransferase and 3,5-epimerase, respectively (Table 2). Since both the megosamine and desosamine pathways require an aminotransferase and a dimethyltransferase, and since mycarose and megosamine each require a 2,3-dehydratase and a 3,5-epimerase, assignments of these four genes to a specific pathway could not be made on the basis of sequence comparison alone. However, the latter three are implicated in megosamine biosynthesis by experiments described below.

*Other genes.* Two additional complete ORFs, designated *megY* and *megH* and an incomplete ORF, designated *megF*, were also identified in the cluster. MegH and MegF share high degrees of similarity with EryH and EryF. EryH and homologs in other macrolide gene clusters are thioesterase-like proteins with unknown function in polyketide gene clusters (Haydock *et al.*, 1991; Xue *et al.*, 1998; Butler *et al.*, 1999; Tang *et al.*, 1999). EryF encodes the erythronolide B C-6 hydroxylase (Figure 8) (Weber *et al.*, 1991; Andersen and Hutchinson, 1992). MegY does not have an *ery* counterpart but appears to belong to a (small) family of *O*-acyltransferases that transfer short acyl chains to macrolides. Two classes exist: AcyA and MidmB transfer acetyl or propionyl groups to the C-3 hydroxyls on 16-membered macrolide rings (Arisawa *et al.*, 1994; Hara and Hutchinson, 1992); CarE and Mpt transfer isovalerate or propionate to the mycarosyl moiety of carbomycin and midecamycin, respectively (Epp *et al.*, 1989; Arisawa *et al.*, 1993;

Gu *et al.*, 1996). The structures of various megalomicins suggest that MegY belongs to the latter class and is the acyltransferase which converts megalomicin A to megalomicins B, C1, or C2 (verified experimentally below).

5 *Heterologous expression of the meg PKS genes.*

The wild type and genetically modified versions of the *ery* DEBS have been used extensively in heterologous *Streptomyces* hosts for enzyme studies and the production of novel polyketide compounds. Given the similarities between the *ery* and *meg* DEBSs, production characteristics were compared in a commonly  
10 used *Streptomyces* host strain. The three *megA* ORFs were cloned into the expression plasmid pKAO127'kan' (Ziermann and Betlach, 1999) in place of the *eryA* ORFs. Both plasmids, pKAO127'kan' encoding *ery* DEBS and pKOS108-06 encoding *meg* DEBS, were introduced in *Streptomyces lividans* K4-114 and the production of 6-dEB was determined in shake-flask fermentations. The production  
15 profiles were similar in both cases and the maximum titer of 6-dEB was between 30-40 mg/L. In addition, both PKSs produced small amounts (~5%) of 8,8a-deoxyoleandolide, which results from the priming of the PKS with acetate instead of propionate (Kao *et al.*, 1994b). This observation indicates that the loading AT domains of the PKSs display similar relaxed specificities towards starter units.

20

*Conversion of erythromycin to megalomicin in S. erythraea.*

An examination of the *meg* cluster revealed that the putative megosamine biosynthetic genes are clustered directly upstream of the PKS genes. If the hypothesis that these genes are sufficient for biosynthesis and attachment of  
25 megosamine to an erythromycin intermediate is correct, then functional expression of these genes in a strain which produces erythromycin, such as *S. erythraea*, should result in production of megalomicin. A 12 kb DNA fragment carrying all the genes between the leftmost *Xho*I site and the *Eco*RI site (Figure 9) was integrated in the chromosome of *S. erythraea* using the site-specific integrating  
30 vector pSET152 (Bierman *et al.*, 1992). It was surmised that the left and right ends of this fragment would contain necessary promoter regions for transcription of the convergent set of genes in *M. megalomicea* and that they would likely operate in *S. erythraea*.

Fermentation broth from *S. erythraea*/KOS97-42, which contains the integrated *meg* genes, was analyzed by LC/MS and compared to LC/MS profiles of the parent *S. erythraea* strain without the *meg* genes, as well as to megalomicin standards purified from *M. megalomicea*. The new strain was found to produce a mixture of erythromycin A and various megalomicins (~4:1 ratio), thereby showing that the predicted megosamine biosynthetic and glycosyltransferase genes are contained within the cloned *meg* fragment. The two most abundant congeners identified were megalomicins B and C1. Megalomicin A and C2 were also detected in smaller amounts. The presence of the megalomicins B, C1 and C2 also provides direct evidence for the function of the *O*-acyl transferase, MegY, which is present in the integrated *meg* fragment.

### Discussion

The homologies observed among modular PKSs enabled the use of *ery* PKS genes to clone the *meg* biosynthetic gene cluster from *M. megalomicea*. The close similarities between the megalomicin and erythromycin biosynthetic pathways is also reflected in the overall organization of their genes and in the high degree of homology of the corresponding individual gene-encoded polypeptides. Production of 6-dEB from *meg* DEBS in *S. lividans* and conversion of erythromycin to megalomicin using the *megD* genes in *S. erythraea* provides direct evidence that the identified gene cluster is responsible for synthesis of megalomicin.

As seen in Figure 9, the ~ 40 kb segments of the two clusters beginning with *ery/megBV* on the left through the *ery/megF* genes retain a nearly identical organizational arrangement. The notable differences in this region are *eryG* and IS1136 which are absent from the segment of the *meg* cluster analyzed. The *eryG* gene encodes an S-adenosylmethionine (SAM)-dependent mycarosyl methyltransferase that converts erythromycin C to erythromycin A (Figure 8) (Weber *et al.*, 1990; Haydock *et al.*, 1991). The mycarose moiety is modified by esterification (MegY) in megalomicin biosynthesis (Figure 8) and, therefore, the absence of an *eryG* homolog would be expected in the *meg* cluster. The IS1136 element located between *eryAI* and *eryAII* (Donadio and Staver, 1993) is not

known to play a role in erythromycin biosynthesis and its origin in the *ery* cluster has not been determined.

Upstream of the common *meg/eryBV* and *BV* genes, the gene clusters diverge. The ~ 6 kb segment between *eryBV* and *eryK*, the left border of the *ery* gene cluster (Pereda *et al.*, 1997), contains the remaining genes required for mycarose (*eryBVI* and *BVII*) and desosamine biosynthesis (*eryCIV*, *CV*, and *CVI*) and the C-12 hydroxylase (*eryK*) (Stassi *et al.*, 1993). In contrast, the region upstream of *megBV* encodes a set of genes (*megDI-DVII* and *megY*) which can account for all the activities unique to megalomicin biosynthesis (Figure 9). Since introduction of this *meg* DNA segment into *S. erythraea* results in production of megalomicins, it is clear that these genes encode the functions for TDP-megosamine biosynthesis and transfer to its putative substrate erythromycin C, and to acylate megalomicin A (Figure 8). The remaining region upstream of *megDVI* should therefore encode genes only for mycarose and desosamine biosynthesis.

Olano *et al.* (Olano *et al.*, 1999) have recently described a pathway for biosynthesis of TDP-L-daunosamine, a deoxysugar component of the antitumor compounds daunorubicin and doxorubicin produced by *Streptomyces peucetius*. Their pathway proposes four steps from the intermediate TDP-4-keto-6-deoxyglucose controlled by the gene cluster *dnmJQTUVZ*, although the functions for *dnmQ* and *dnmZ* could not be identified and the precise order of reactions in the pathway could not be determined. The genes *dnmT*, *dnmU*, *dnmJ* and *dnmV* each have proposed counterparts in the *meg* cluster, *megT*, *megDIV*, *megDII*, and *megDV*, respectively (see Figure 10)

It is possible to describe a pathway to convert TDP-2,6-dideoxy-3,4-diketo-D-hexose (or its enol tautomer), the last intermediate common to the mycarose and megosamine pathways, to TDP-megosamine through the sequence of 5-epimerization, 4-ketoreduction, 3-amination, and 3-*N*-dimethylation employing the genes *megDIV*, *megDV*, *megDII*, and *megDIII*. This employs the same functions proposed for biosynthesis of TDP-daunosamine by Olano *et al.*, but in a different sequential order. However, it does not account for the *megDVI* and *megDVII* genes since their activities are not required for this route. A parallel pathway which employs these genes is also shown in Figure 10. In this alternate route, 2,3-reduction and 3,4-tautomerization are performed by the *megDVII* and

*megDVI* gene products, respectively. A unified single pathway that employs both 4-ketoreduction (*megDV*) and 2,3-reduction (*megDVII*) could not be determined. Because the entire gene set from *megDVI* through *megDVII* was introduced in *S. erythraea* to produce TDP-megosamine, it is not possible to determine which, if  
5 either, of the two alternative pathways is operative, but this can be addressed through systematic gene disruption and complementation.

The 48 kb segment sequenced also contains genes required for synthesis of TDP-L-mycarose and TDP-D-desosamine (Fig 10). For the latter, *megCII*, which encodes a putative 3,4-isomerase, the first step in the committed TDP-desosamine  
10 pathway, appears to be translationally coupled to *megAIII*, almost exactly as its erythromycin counterpart, *eryCII*, was found translationally coupled to *eryAIII* (Summers *et al.*, 1997). The high degree of similarity between MegCII and EryCII suggests that the pathway to desosamine in the megalomicin- and erythromycin-producing organisms are most likely the same. Similarly, the finding that *megBII*  
15 and *megBIV*, encoding a 2,3-reductase and 4-ketoreductase, contain close homologs in the mycarose pathway for erythromycin also suggests that TDP-L-mycarose synthesis in the two host organisms is the same.

Of interest are the two genes that encode putative 2,3-reductases, *megBII* and *megDVII*. Because MegBII most closely resembles EryBII, a known mycarose  
20 biosynthetic enzyme (Weber *et al.*, 1990), and because *megBII* resides in the same location of the *meg* cluster as its counterpart in the *ery* cluster, *megBII* is assigned to the mycarose pathway and *megDVII* to the megosamine pathway. Furthermore, the lower degree of similarity between MegDVII and either EryBII or MegBII (Table 2) provides a basis for assigning the opposite L and D isomeric substrates  
25 to each of the enzymes (Figure 10). Finally, *megT*, which encodes a putative 2,3-dehydratase, is also related to a gene in the *ery* mycarose pathway, *eryBVI*. In *S. erythraea*, the proposed intermediate generated by EryBVI represents the first committed step in the biosynthesis of mycarose (Figure 10). However, the proposed pathways in Figure 10 suggest this may be an intermediate common to  
30 both mycarose and megosamine biosynthesis in *M. megalomicea*. Therefore, *megT* is named following the designation of the equivalent gene in the daunosamine pathway, *dnmT* (Olano *et al.*, 1999)

The preferred host-vector system for expression of *meg* DEBS described here has been used previously for the heterologous expression of modular PKS genes from the erythromycin (Kao *et al.*, 1994a; Ziermann and Betlach, 1999), picromycin (Tang *et al.*, 1999) and oleandomycin pathways, as well as for the  
5 generation of novel polyketide backbones where domains have been removed, added or exchanged in various combinations (McDaniel *et al.*, 1999). Recently, hybrid polyketides have been generated through the co-expression of subunits from different PKS systems (Tang *et al.*, 2000).

Expression of the *megDVI-megDVII* segment in *S. erythraea* and the  
10 corresponding production of megalomicins in this host establishes the likely order of sugar attachment in megalomicin synthesis. Furthermore, it provides a means to produce megalomicin in a more genetically friendly host organism, leading to the creation of megalomicin analogs by manipulating the PKS. Over 60 6-dEB analogs have been produced by combinatorial biosynthesis using the *ery* PKS  
15 (McDaniel *et al.*, 1999; Xue *et al.*, 1999). The titers of megalomicin could also be significantly increased above the 5 mg/L obtained from *M. megalomiciea* by introducing the genes into an industrially optimized strain of *S. erythraea*, many of which can produce as much as 10 g/L of erythromycin.

## 20 References

- Kao, C.M., Katz, L. and Khosla, C. (1994a) Engineered biosynthesis of a complete macrolactone in a heterologous host. *Science* **265**: 509-512.
- Kao, C.M., Luo, G., Katz, L., Cane, D.E. and Khosla, C. (1994b) Engineered  
25 biosynthesis of a triketide lactone from an incomplete modular polyketide synthase. *J. Am. Chem. Soc.* **116**: 11612-11613.
- McDaniel, R., Thamchaipenet, A., Gustafsson, C., Fu, H., Betlach, M., Betlach, M. *et al.* (1999) Multiple genetic modifications of the erythromycin gene cluster to produce a library of novel "unnatural" natural products. *Proc.*  
30 *Natl. Acad. Sci. USA* **96**: 1846-1851.
- Olano, C., Lomovskaya, N., Fonstein, L., Roll, J.T. and Hutchinson, C.R. (1999) A two-plasmid system for the glycosylation of polyketide antibiotics:



- bioconversion of e-rhodomyacinone to rhodomyacin D. *Chem. & Biol.* 6: 845-855.
- Tang, L., Fu, H., Betlach, M.C. and McDaniel, R. (1999) Elucidating the mechanism of chain termination switching in the picromycin/methymycin polyketide synthase. *Chem. & Biol.* 6: 553-558.
- Tang, L., Fu, H. and McDaniel, R. (2000) Formation of functional heterologous complexes using subunits from the picromycin, erythromycin, and oleandomycin polyketide synthases. *Chem. & Biol.* 7: 77-84.
- Weber, J.M., Leung, J.O., Maine, G.T., Potenz, R.H., Paulus, T.J. and DeWitt, J.P. (1990) Organization of a cluster of erythromycin genes in *Saccharopolyspora erythraea*. *J. Bacteriol.* 172: 2372-2383.
- Weber, J.M., Leung, J.O., Swanson, S.J., Idler, K.B. and McAlpine, J.B. (1991) An erythromycin derivative produced by targeted gene disruption in *Saccharopolyspora erythraea*. *Science* 252: 114-117.
- Xue, Q., Ashley, G., Hutchinson, C.R. and Santi, D.V. (1999) A multi-plasmid approach to preparing large libraries of polyketides. *Proc. Natl. Acad. Sci. USA* 96: 11740-11745.
- Xue, Y., Zhao, L., Liu, H.-w. and Sherman, D.H. (1998) A gene cluster for the macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity. *Proc. Natl. Acad. Sci. USA* 95: 12111-12116.
- Ziermann, R. and Betlach, M. (2000) A two-vector system for the production of recombinant polyketides in *Streptomyces*. *J. Ind. Microbiol. Biotech.* 24: 46-50.
- Ziermann, R. and Betlach, M.C. (1999) Recombinant polyketide synthesis in *Streptomyces*: Engineering of improved host strains. *Biotechniques* 26: 106-110.

## Example 2

### Stabilizing meg PKS Expression Plasmid by Codon Engineering

#### 30 *Materials and methods*

All bacterial strains were cultured and transformed as described in Example 1.

*Fermentation of Streptomyces and diketide feeding*

Primary *Streptomyces* transformants were picked and placed in 6 mL of TSB liquid medium with 50 µg/L of thiostrepton and grown at 30°C. When the culture showed some growth (3-4days), it was transferred into a 250 mL flask containing 50 mL of R6 medium (pH 7.0) with 25 ug/L of thiostrepton and 1g/L of diketide ((2s,3R)2-methyl-3-hydroxyhexanoate N-propionyl cysteamine thioester) and placed in a 30°C incubator for 7 days.

10 *Changing codons and making plasmids*

There are several identical sequences in the coding sequences for module 2 and module 6 of the megalomicin PKS gene cluster. Expression plasmids containing the full length megalomicin PKS appeared to be somewhat unstable and subject to deletion in *recA*<sup>+</sup> strains like ET124567 and *Streptomyces* by intra-plasmid homologous recombination. To prevent significant homologous recombination and so stabilize expression plasmids, the codons of two regions of the module 6 coding sequence that are identical to regions in the module 2 coding sequence were changed without changing the sequence of protein encoded. The two regions changed in module 6 were from the 26739<sup>th</sup> base to 27,267<sup>th</sup> base and from position 27,697<sup>th</sup> base to 27,987<sup>th</sup> base, which were identical to the region from position 6810<sup>th</sup> base to 7338<sup>th</sup> base and regions from position 7778<sup>th</sup> base to 8068<sup>th</sup> base, respectively. The start codon of the loading domain of the meg PKS was set to be the 1<sup>st</sup> base. These sequences are shown below

25 > 6810-7338 Sequence in Module 2  
 TTGCAGCGGTTGTCGGTGGCGGTGCGGGAGGGGCGTCCGGTGTGGGTGTGGTGGTGGGT  
 TCGGCGGTGAATCAGGATGGGGCGAGTAATGGGTTGGCGGCGCCGTCCGGGGTGGCGCAG  
 CAGCGGGTGATTCGGCGGGCGTGGGGTTCGTGCGGGTGTGTCGGGTGGGGATGTGGGTGTG  
 GTGGAGGCGCATGGGACGGGGACGCGGTTGGGGGATCCGGTGGAGTTGGGGGCGTTGTTG  
 30 GGGACGTATGGGGTGGGTCCGGGTGGGGTGGGTCCGGTGGTGGGTTCGGTGAAGGCG  
 AATGTGGGTTCATGTGCAGGCGGCGGGGTGTGGTGGGTGTGATCAAGGTGGTGTGGGG  
 TTGGGTCCGGGGTTGGTGGGTCCGATGGTGTGTCGGGGTGGGTTCGGGGTTGGTGGAT  
 TGGTCGTCGGGTGGGTGGTGGTGGCGGATGGGGTGGGGGGTGGCCGGTGGGTGTGGAT  
 GGGGTGCGTCCGGGTGGGGTGTCCGCGTTTGGGGTGTCCGGGACGAAT (SEQ ID NO: 23)  
 35 > 26736-27267 Sequence in Module 6  
 CTGCAGCGGTTGTCGGTGGCGGTGCGGGAGGGGCGTCCGGTGTGGGTGTGGTGGTGGGT  
 TCGGCGGTGAATCAGGATGGGGCGAGTAATGGGTTGGCGGCGCCGTCCGGGGTGGCGCAG  
 CAGCGGGTGATTCGGCGGGCGTGGGGTTCGTGCGGGTGTGTCGGGTGGGGATGTGGGTGTG  
 GTGGAGGCGCATGGGACGGGGACGCGGTTGGGGGATCCGGTGGAGTTGGGGGCGTTGTTG  
 40 GGGACGTATGGGGTGGGTCCGGGTGGGGTGGGTCCGGTGGTGGGTTCGGTGAAGGCG  
 AATGTGGGTTCATGTGCAGGCGGCGGGGTGTGGTGGGTGTGATCAAGGTGGTGTGGGG

TTGGGTCGGGGGTTGGTGGGTCCGATGGTGTGTCGGGGTGGGTGTCGGGGTGGTGGAT  
 TGGTCGTCGGGTGGGTGGTGGTGGCGGATGGGGTGCGGGGGTGGCCGGTGGGTGTGGAT  
 GGGGTGCGTCGGGGTGGGGTGTGCGGCGTTTGGGGTGTGCGGGGACGAAT (SEQ ID NO: 24)  
 > 26736-27267 Sequence with Codon Changes  
 5 CTGCAGCGCCTCTCCGTCGCGGTCCGCGAGGGCCGCGAGTCCTCGGCGTCGTCGTCGGC  
 TCGGCCGTCAACCAAGACGGCGCGTCAAACGGCCTCGCCGCGCCCTCCGGCGTCGCCCAG  
 CAGCGCGTCATACGCCGCGCGTGGGGACGCGCCGAGTATCGGGCGGCGACGTCGGAGTC  
 GTCGAGGCCCCACGGCACCGGCACCCGCTCGGGGATCCCGTCGAGCTGGGCGCCCTCCTG  
 GGCACGTACGGCGTCGGCCGCGGCGGCGTCTCGGCCGGTCGTCGTCGGCAGCGTCAAGGCC  
 10 AACGTCGGCCACGTCCAGGCCGCGGCGGCGTCTCGGGGTCAATCAAGGTCGTCCTCGGC  
 CTCGGCCGCGGGCTGGTCGGCCCGATGGTCTGCCGCGGCGGCGCTCAGCGGCCCTCGTCGAC  
 TGGTCGTCCGGCGGCGCTGGTCGTCGCGGACGGGGTCCGCGGCTGGCCGGTCGGCGTCGAC  
 GCGTCCGCGGGGCGGCGTCTCGGCGTTCGGCGTCAGCGGGACGAAT (SEQ ID NO: 25)

15 > 6978-7337 Sequence in Module 2  
 GGTGGAGTGTGATGCGGTGGTGTGTCGTCGGTGGTGGGGTTTTTCGGTGTGGGGGTGTTGGA  
 GGGTCGGTCGGGTGCGCCGTCGTTGGATCGGGTGGATGTGGTGCAGCCGGTGTGTTTCGT  
 GGTGATGGTGTGCTTGGCGCGGTTGTGGCGGTGGTGTGGGGTGTGCCTGCGGCGGTGGT  
 20 GGGTCATTCGCAGGGGGAGATCGCGGCGGCGGTGGTGGCGGGGGTGTGTCGGTGGGTGA  
 TGGTGC GCGGGTGGTGGCGTTGCGGGCGCGGGCGTTGCGGGCGTTGGCCGG (SEQ ID NO:  
 26)  
 > 27697-27987 Sequence in Module 6  
 GGTGGAGTGTGATGCGGTGGTGTGTCGTCGGTGGTGGGGTTTTTCGGTGTGGGGGTGTTGGA  
 GGGTCGGTCGGGTGCGCCGTCGTTGGATCGGGTGGATGTGGTGCAGCCGGTGTGTTTCGT  
 25 GGTGATGGTGTGCTTGGCGCGGTTGTGGCGGTGGTGTGGGGTGTGCCTGCGGCGGTGGT  
 GGGTCATTCGCAGGGGGAGATCGCGGCGGCGGTGGTGGCGGGGGTGTGTCGGTGGGTGA  
 TGGTGC GCGGGTGGTGGCGTTGCGGGCGCGGGCGTTGCGGGCGTTGGCCGG (SEQ ID NO:  
 27)  
 > 27697-27987 Sequence with Codon Changes  
 30 CGTGGAGTGC GATGCGGTGCTGTCGAGCGTCGTCGGCTTCAGCGTGCTGGGCGTCCTGGA  
 GGGCCG CAGCGGCGCCCGAGCCTGGACCGCGTCGACGTGGTCCAGCCGGTCCTGTTTCGT  
 GGTCATGGTCAGCCTGGCCCGCCTGTGGCGCTGGTGC GCGGTGGTCCCGGCCGCGGTGGT  
 CCGCCACAGCCAGGGCGAGATCGCCGCCGCGGTGCTGGCCGGCGTCTGAGCGTCGGCGA  
 35 CCGCGCCCGCGTCGTGGCCCTGCGCGCCCGCGCCCTGCGCGCCCTGGCCGG (SEQ ID NO:  
 28)

Three pieces of DNA from the two regions above were synthesized and verified by  
 Retrogen, and the synthesized DNAs were cloned into pCR-Blunt II -TOPO, as  
 shown in the Table 3 below.

Table 3. Plasmids containing synthesized DNA

Plasmids	Cloning sites and positions in meg PKS
pKOS97-1613	PstI-BamHI, 26,739 <sup>th</sup> -26,947 <sup>th</sup> base
PKOS97-1622	BamHI-BsmI, 26,947 <sup>th</sup> -27,267 <sup>th</sup> base
PKOS97-1628	SfaNI-FseI, 27,697 <sup>th</sup> -27,987 <sup>th</sup> base

#### *Assembly of the expression plasmid*

First, ligation of the PstI-BamHI fragment of pKOS97-1613, the BamHI-  
 45 BsmI fragment of pKOS97-1622 and BsmI-PstI linearized pKOS97-90 produced

pKOS97-151. Then, the insertion of the SfaNI-FseI fragment of pKOS97-1628 into pKOS97-151 gave rise to pKOS97-152. Then, the PstI-BlnI fragment of pKOS97-125 was used to replace the PstI-BlnI fragment of pKOS97-90a and produced pKOS97-160.

5           The final expression plasmid (in pRM5) pKOS97-162 was the result of BglII-NheI fragment of pKOS97-160 inserted into BglII-NheI sites of pKOS108-04.

Another expression plasmid pKOS97-152a was made by a four-fragment ligation. The four fragments were a BlnI-XbaI fragment (containing a cos site) of  
10   pKOS97-92a, a BglII-PstI fragment of pKOS97-81, a PstI-BlnI fragment of pKOS97-152, and a BglII-XbaI fragment of pKOS108-04 (as the vector).

Tests of the constructed plasmids showed that the plasmids containing the modified coding sequences were more stable than plasmids containing unmodified coding sequence.

15

### Example 3

#### Construction of Ole-Meg Hybrid PKS

*Construction of pRM1-based pKOS98-48 for the expression of OlePKS modules 1-4.*

20           The 240-bp fragment containing the 3'-end portion of *oleAII* gene (at nt 11210-11452; the first base of the start codon of *oleAII* is nt 1) was PCR amplified with primers N98-38-1 (5'-GAACAACCTCCTGTCTGCGGCCGCG-3') (SEQ ID NO: 29) and N98-38-3 (5'-  
CGGAATTCTCTAGAGTCACGTCTCCAACCGCTTGTCGAGG-3') (SEQ ID  
25 NO: 30). The fragment contains a naturally occurring NotI site at its 5'-end and the engineered XbaI (bold) and EcoRI sites (underline) at its 3'-end following the *oleAII* stop codon. pKOS38-189 was digested with EcoRI and NotI to give five fragments of 8-kb, 5 kb, 4 kb, 2.5 kb and 2 kb. The 8-kb EcoRI-NotI fragment containing *oleAII* gene nt 2961 to nt 11210 and the 240-bp NotI, EcoRI treated  
30 PCR fragment were ligated into litmus 28 at the EcoRI site via a three-fragment ligation to give pKOS98-46. The 8.2-kb EcoRI fragment from pKOS98-46 was cloned into pKOS38-174, a pRM1 derived plasmid containing *oleAI* and nt 1 to nt 2960 of *oleAII* to give pKOS98-48.

*Construction of pSET152-based pKOS98-60 for the expression of megPKS modules 5-6.*

The 360-bp fragment containing nt 1 to nt 366 of *megAIII* was PCR  
5 amplified with primers N98-40-3 (5'-  
TCTAGACTTAATTAAGGAGGACACATATGAGCGA-GAGCAGC-  
GGCATGACCG-3') (SEQ ID NO: 31) and N98-40-2 (5'- AACGCCTCCCAG-  
GAGATCTCCAGCA-3') (SEQ ID NO: 32). A *PacI* site and a *NdeI* site as well  
as the ribosome binding site were introduced at the 5'-end of the *megAI* start  
10 codon. The 360-bp *PacI*-*BglII* fragment was inserted into pKOS108-06 replacing  
the 22-kb *PacI*-*BglII* fragment to yield pKOS98-55. The 10-kb *PacI*-*XbaI*  
fragment containing *megAIII* gene and the annealed oligos N98-23-1 (5'-  
AATTCATAGCCTAGGT-3') (SEQ ID NO: 33) and N98-23-2 (5'-  
CTAGACCTAGGCTATG-3') (SEQ ID NO: 34) were ligated to *PacI* and *EcoRI*  
15 treated pSET152 derivative pKOS98-14 via a three-fragment ligation to give  
pKOS98-60.

Example 4

Conversion of Erythronolides to Erythromycins

20 A sample of a polyketide (~50 to 100 mg) is dissolved in 0.6 mL of  
ethanol and diluted to 3 mL with sterile water. This solution is used to overlay a  
three day old culture of *Saccharopolyspora erythraea* WHM34 (an *eryA* mutant)  
grown on a 100 mm R2YE agar plate at 30°C. After drying, the plate is incubated  
at 30°C for four days. The agar is chopped and then extracted three times with 100  
25 mL portions of 1% triethylamine in ethyl acetate. The extracts are combined and  
evaporated. The crude product is purified by preparative HPLC (C-18 reversed  
phase, water-acetonitrile gradient containing 1% acetic acid). Fractions are  
analyzed by mass spectrometry, and those containing pure compound are pooled,  
neutralized with triethylamine, and evaporated to a syrup. The syrup is dissolved  
30 in water and extracted three times with equal volumes of ethyl acetate. The  
organic extracts are combined, washed once with saturated aqueous NaHCO<sub>3</sub>,  
dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, and evaporated to yield ~0.15 mg of product. The  
product is a glycosylated and hydroxylated compound corresponding to

erythromycin A, B, C, and D but differing therefrom as the compound provided differed from 6-dEB.

#### Example 5

5

#### Measurement of Antibacterial Activity

Antibacterial activity is determined using either disk diffusion assays with *Bacillus cereus* as the test organism or by measurement of minimum inhibitory concentrations (MIC) in liquid culture against sensitive and resistant strains of *Staphylococcus pneumoniae*.

10

#### Example 6

#### Evaluation of Antiparasitic Activity

Compounds can initially be screened *in vitro* using cultures of *P. falciparum* FCR-3 and K1 strains, then *in vivo* using mice infected with *P. berghei*. Mammalian cell toxicity can be determined in FM3A or KB cells. Compounds can also be screened for activity against *P. berhei*. Compounds are also tested in animal studies and clinical trials to test the antiparasitic activity broadly (antimalarial, trypanosomiasis and Leishmaniasis).

15

The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments and that the foregoing description and examples are for purposes of illustration and not limitation of the following claims.

25



Claims

1. An isolated nucleic acid comprising a nucleotide sequence encoding a domain of megalomicin polyketide synthase (PKS) or a megalomicin modification enzyme.  
5
2. The isolated nucleic acid of claim 1, which encodes a PKS open reading frame (ORF) selected from the group consisting of megAI, megAII and megAIII.
- 10 3. The isolated nucleic acid of claim 1, wherein the PKS domain is selected from the group consisting of a TE domain, a KS domain, an AT domain, an ACP domain, a KR domain, a DH domain, and an ER domain.
4. The isolated nucleic acid of claim 1, wherein the nucleic acid  
15 comprises the coding sequence for a loading module, a thioesterase domain, and all six extender modules of megalomicin PKS.
5. The isolated nucleic acid of claim 1, which encodes a megalomicin modification enzyme that is involved in the conversion of 6-dEB into a  
20 megalomicin.
6. The isolated nucleic acid of claim 5, which encodes a megalomicin modification enzyme that is involved in the biosynthesis of mycarose, megosamine or desosamine.  
25
7. The isolated nucleic acid of claim 1, wherein the nucleic acid codons of homologous regions within the PKS or the megalomicin modification enzyme coding sequence have been changed to reduce or abolish the homology without changing the amino acid sequences encoded by said changed nucleic acid  
30 codons.

8. The isolated nucleic acid of claim 1, which isolated nucleic acid fragment hybridizes to a nucleic acid having a nucleotide sequence set forth in the SEQ. ID NO:1.

5 9. A polypeptide, which is encoded by the isolated nucleic acid fragment of claim 1.

10. A recombinant DNA expression vector, comprising the isolated nucleic acid of claim 1 operably linked to a promoter.

10

11. A recombinant host cell, comprising the recombinant DNA expression vector of claim 10.

12. The recombinant host cell of claim 11, which is a *Streptomyces* or  
15 *Saccharopolyspora* host cell.

13. A recombinant host cell of claim 11, which comprises:

a) at least two separate autonomously replicating recombinant DNA expression vectors, each of said vectors comprises a recombinant DNA compound  
20 encoding a megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter; or

b) at least one autonomously replicating recombinant DNA expression vector and at least one modified chromosome, each of said vector(s) and each of said modified chromosome comprises a recombinant DNA compound encoding a  
25 megalomicin PKS domain or a megalomicin modification enzyme operably linked to a promoter.

14. A hybrid PKS that comprises a polypeptide of claim 9 and is composed of at least a portion of a megalomicin PKS and at least a portion of a  
30 second PKS for a polyketide other than megalomicin.

15. The hybrid PKS of claim 14, wherein the second PKS is selected from the group consisting of a narbonolide PKS, an oleandolide PKS, and a DEBS PKS.

5 16. The hybrid PKS of claim 15 that is composed of the megAI and megAII gene products and the oleAIII gene product.

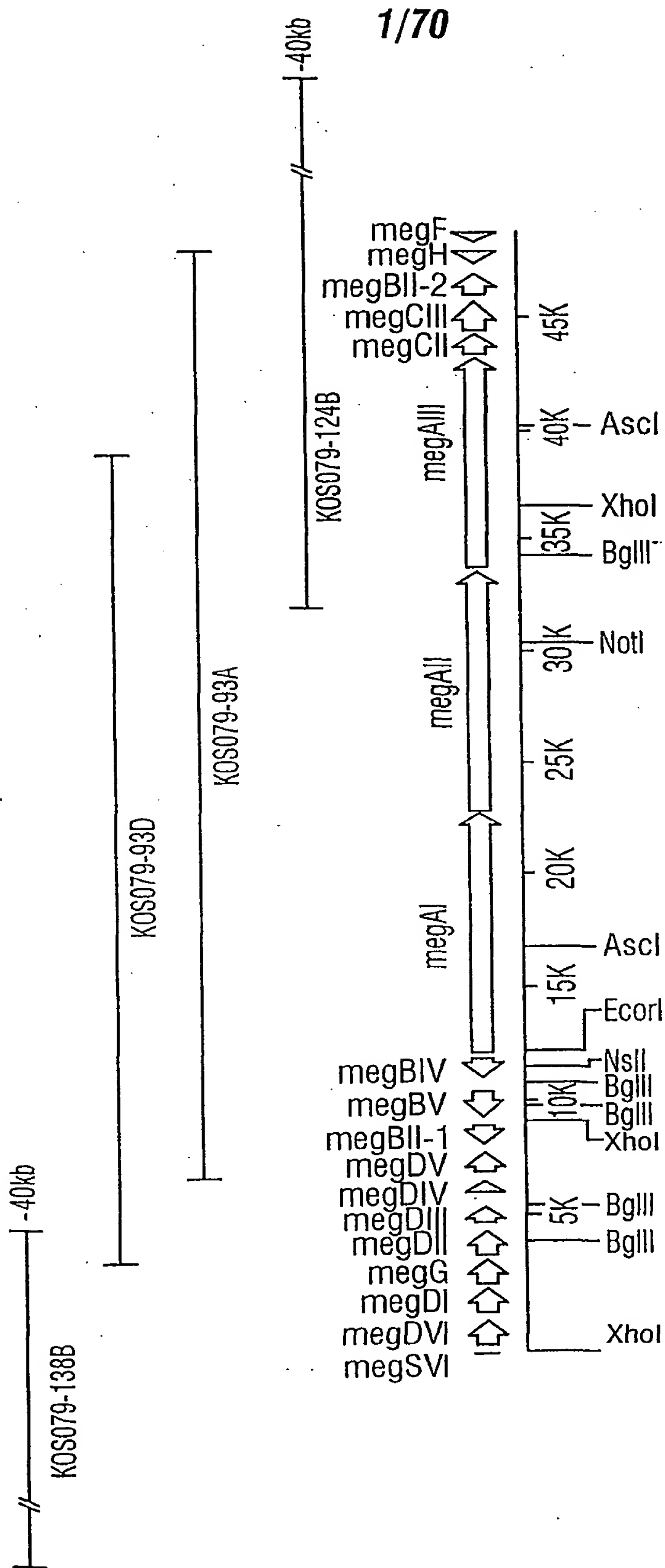
17. The hybrid PKS of claim 16, wherein the KS domain of module 1 of the megAI gene product has been inactivated by mutation.

10

18. A method of producing a polyketide, which method comprises growing the recombinant host cell of claim 11 under conditions whereby the megalomicin PKS domain encoded by the recombinant expression vector is produced and the polyketide is synthesized by the cell, and recovering the  
15 synthesized polyketide.

19. A recombinant host cell that comprises a recombinant expression vector that encodes a megalomicin modification enzyme.

20 20. The recombinant host cell of claim 19 that produces megosamine and can attach megosamine to a polyketide, wherein said host cell, in its naturally occurring non-recombinant state cannot produce megosamine.



**FIG. 1**

2/70

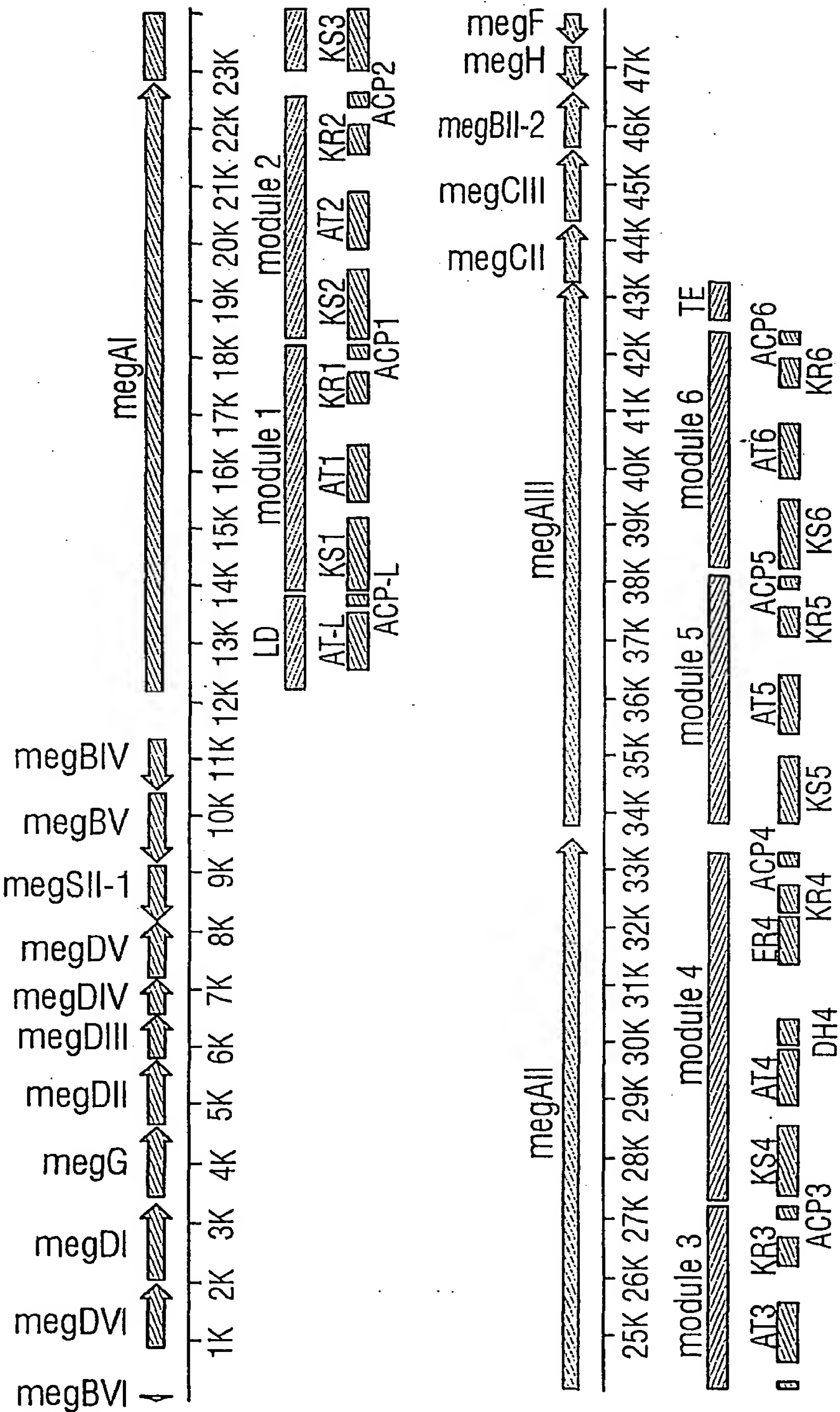
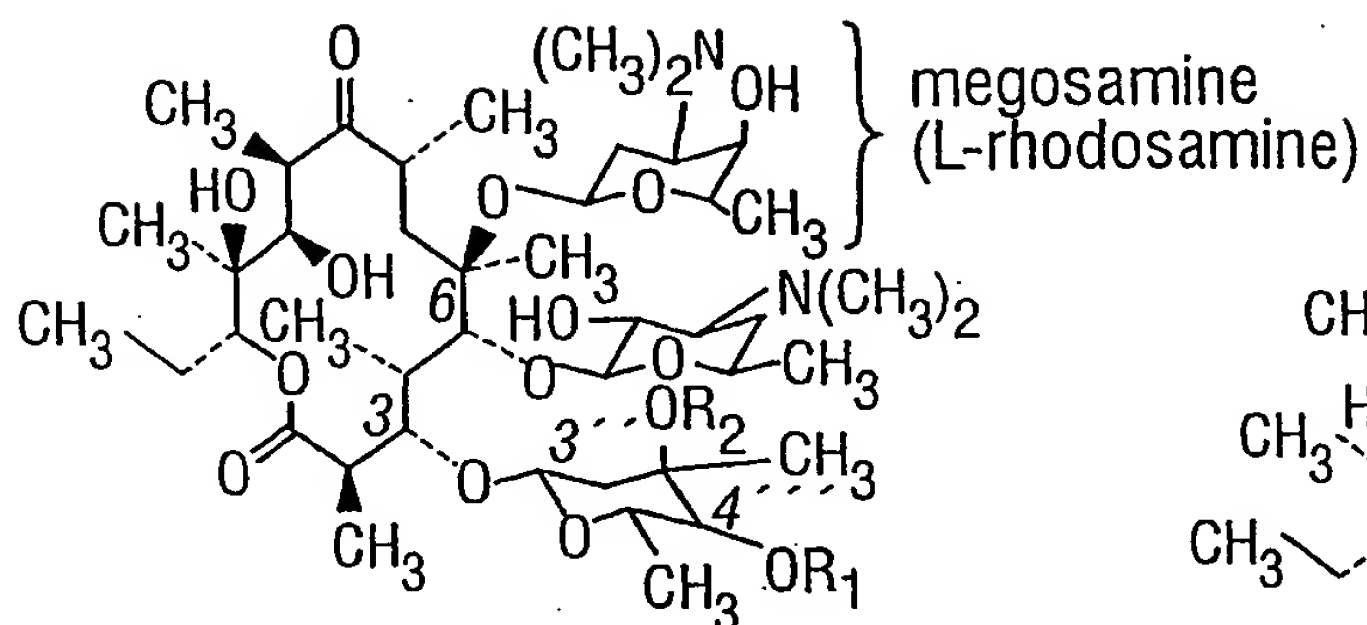


FIG. 2

3/70



		$R_1$	$R_2$
Megalomicin	A	H	H
	B	COCH <sub>3</sub>	H
	C1	COCH <sub>3</sub>	COCH <sub>3</sub>
	C2	COCH <sub>2</sub> CH <sub>3</sub>	COCH <sub>3</sub>

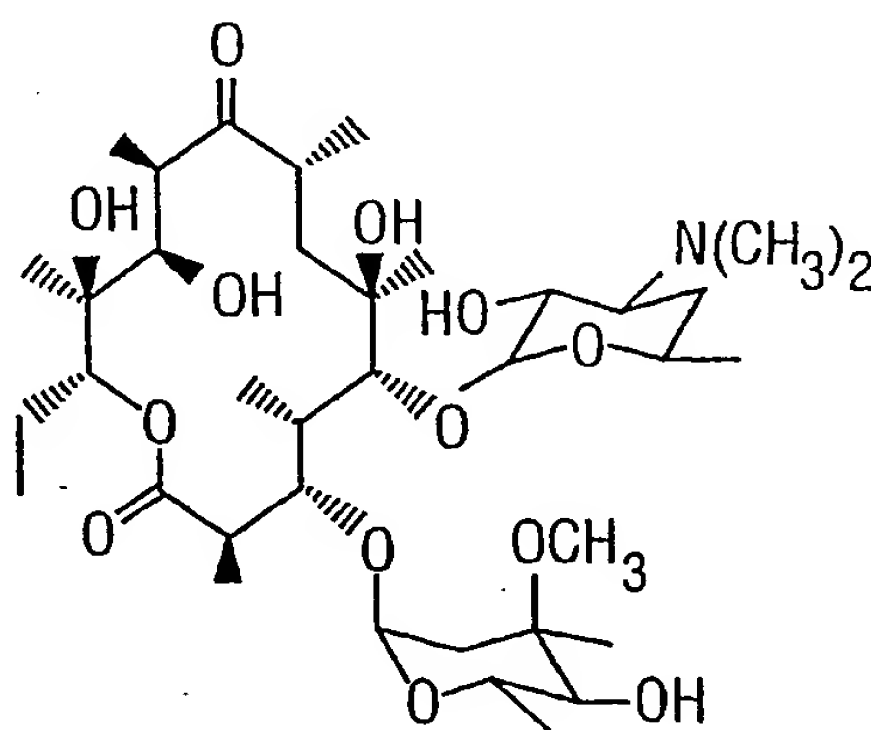
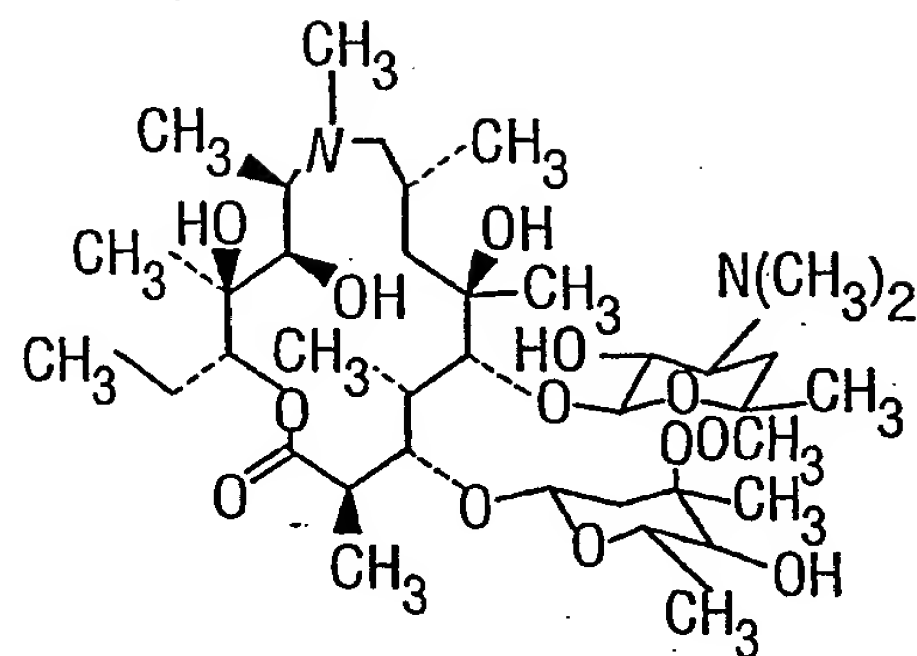


FIG. 3



4/70

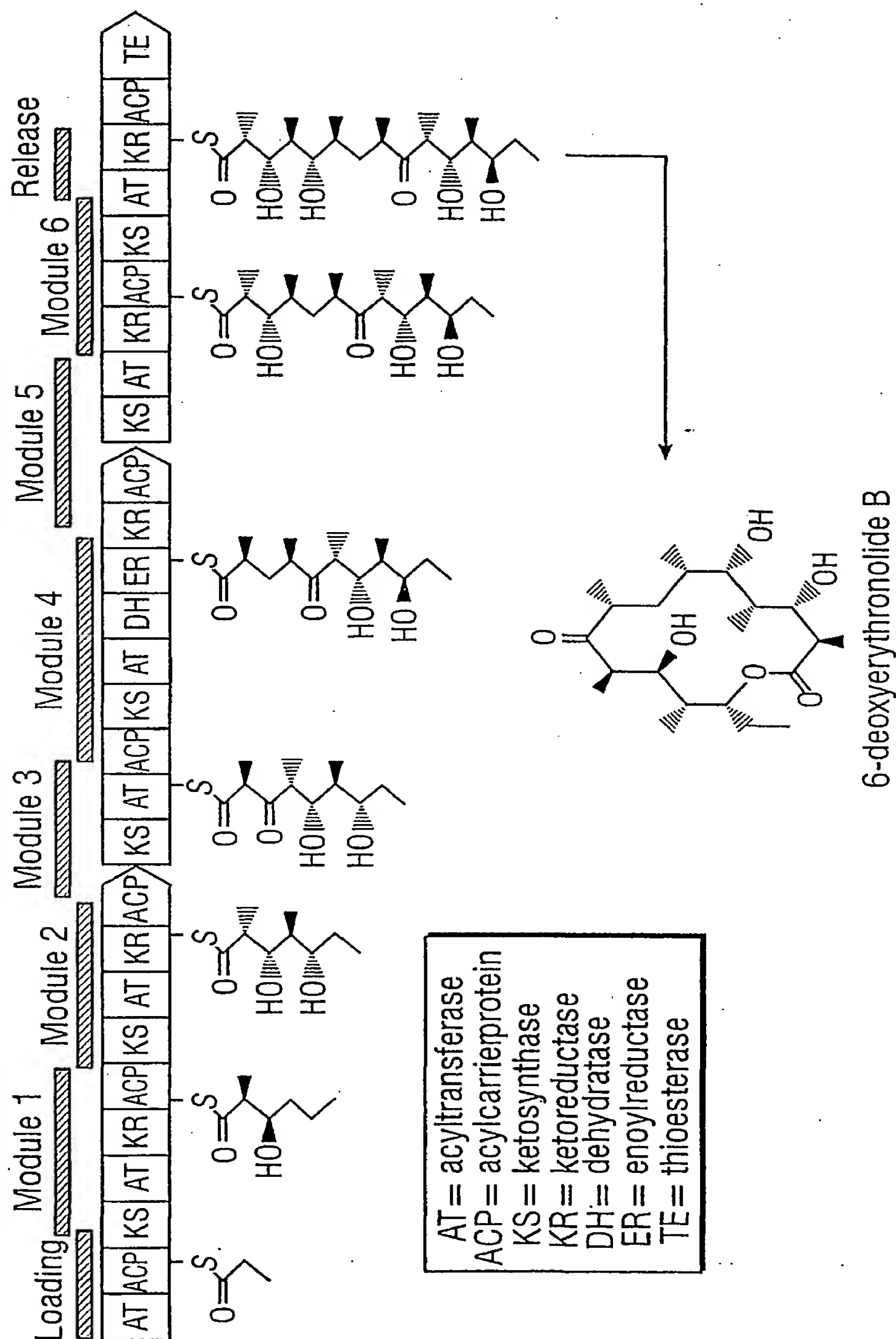


FIG. 4

5/70

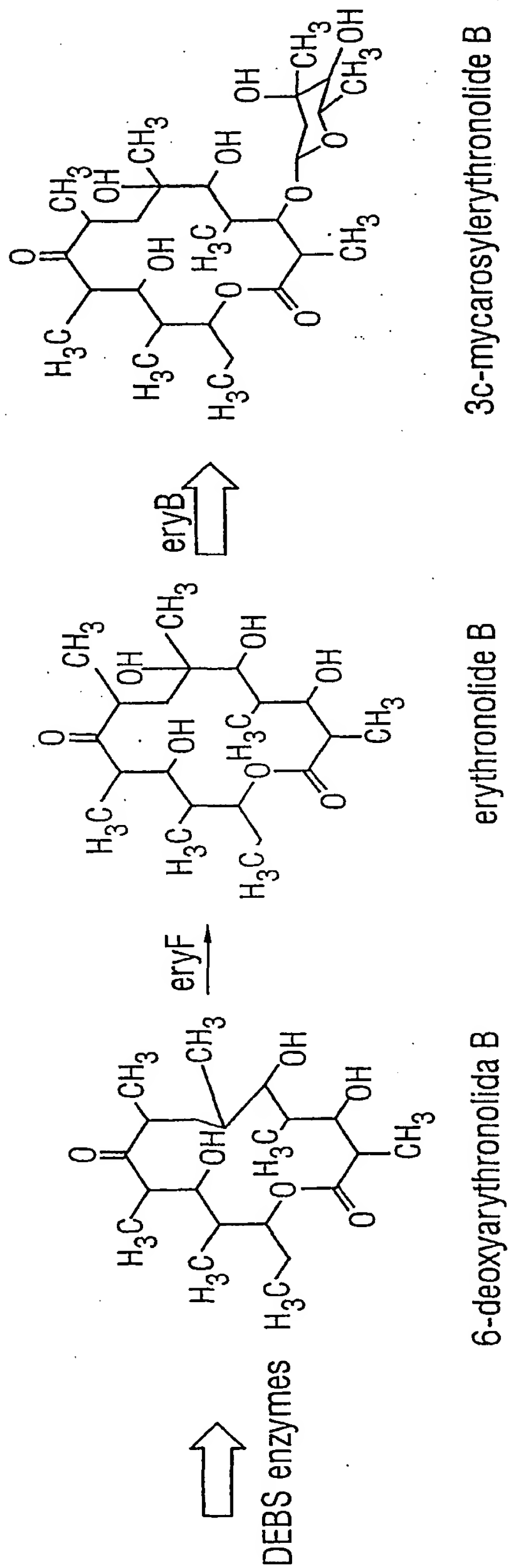


FIG. 5A

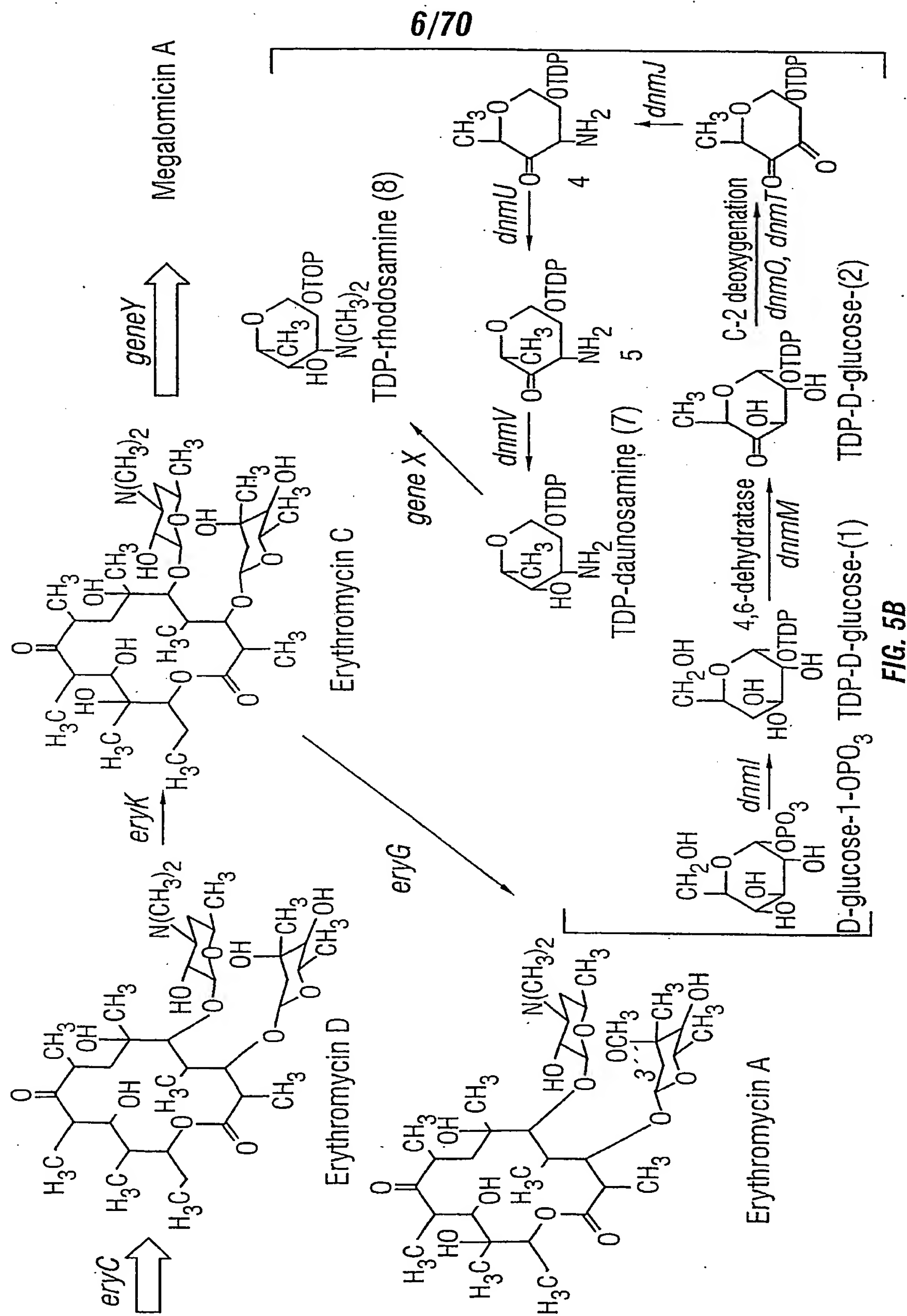


FIG. 5B

7/70

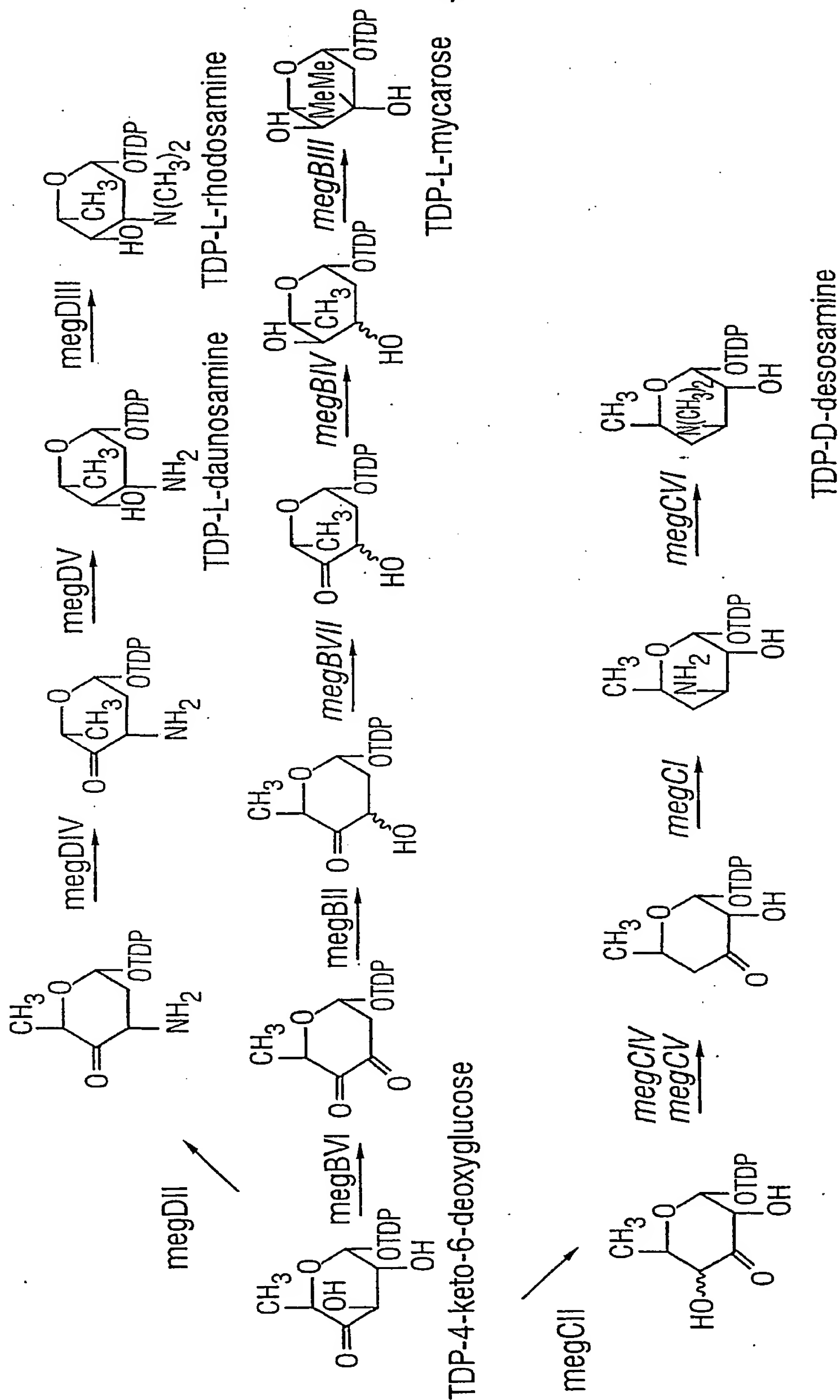


FIG. 6

LOCUS 1 47981 bp DNA 01-MAY-2000

DEFINITION Megalomycin biosynthetic gene cluster, polyketide synthase, desosamine, megosamine, and mycarose biosynthesis genes.

ACCESSION 1

VERSION

KEYWORDS

SOURCE Micromonospora megalomicea.

ORGANISM Micromonospora megalomicea

Unclassified.

REFERENCE 1 (bases 1 to 47981)

AUTHORS Volchegursky, Y., Hu, Z., Katz, L. and McDaniel, R.

TITLE Biosynthesis of the Anti-Parasitic Agent Megalomycin: Transformation of Erythromycin to Megalomycin in *Sacharopolyspora erythraea*

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 47981)

AUTHORS McDaniel, R. and Volchegursky, Y.

TITLE Direct Submission

JOURNAL Submitted (01-MAY-2000) Kosan Biosciences, Inc., 3828 Bay Center Place, Hayward, CA 94545, USA

FEATURES Location/Qualifiers

source 1..47981

/organism="Micromonospora megalomicea"

8/70

FIG. 7-1

9/70

```
gene /strain="NRRL3275"  
/sub_species="nigra"  
complement (<1..144)  
/gene="megT"  
CDS complement (<1..144)  
/gene="megT"  
/codon_start=1  
/transl_table=11  
/product="TDP-4-keto-6-deoxyglucose-2,3-dehydratase"  
/translation="MGDRVNGHATPESTQSAIRFLTRHGGPPTATDDVHDLAHRAAE  
HRLE" (SEQ ID NO: 2)  
gene 928..2061  
/gene="megDVI"  
CDS 928..2061  
/gene="megDVI"  
/codon_start=1  
/transl_table=11  
/product="TDP-4-keto-6-deoxyhexose 3,4-isomerase"  
/translation="MAVGDRRRRLGRELQMARGLYWGFGANGDLYSMLLSGRDDDPWTW  
YERLRAAGRGPYASRAGTWVVGDRHTAEVLADPGFTHGPPDAARWMQVAHCPAASWA  
GPFREFYARTEDAASVTVDADWLQQRCARLVTELGSRFDLVNDFAREVPVLALGTAPA  
LKGVDPRRLRSWTSATRVCLDAQVSPQQLA VTEQALTALDEIDAVTGGRDAAVLGVV  
AELAAANTVGNAVLA VTELP ELAARLADDPETATRVVTEVSRTSPGVHLERRTAASDRR
```

FIG. 7-2



10/70

```

VGGVDVPTGGEVTVVAAANRDPEVFTDPDRFDVDRGGDAEILSSRPGSPRTDLDALV
ATLATAALRAAAPVLPRLSRSGPVIRRRRSPVARGLSRCPVEL" (SEQ ID NO: 3)
gene 2072..3382
      /gene="megDI"
CDS   2072..3382
      /gene="megDI"
      /codon_start=1
      /transl_table=11
      /product="TDP-megosamine glycosyltransferase"
      /translation="MRVVFSSMAVNSHLFGLVPLASAFQAAGHEVRVVASPALTDDVT
GAGLTAVPVGDDVELVEWHAHAGQDIVEYMRITLDWVDQSHTTMSWDDLLGMQTTFTPT
FFALMSPDSLIDGMVEFCRSWRPDWIVWEPLTFAAPIAARVTGTPHARMLWGPDVATR
ARQSFRLRLAHQEVEHREDPLAEWFDTLRRFGDDPHLSFDEELVLGQWTVDPPIPEPL
RIDTGVRTVGMRYVPYNGPSVVPAWLLREPERRRVCCLTLGSSSREHGIGQVSIGEMLD
AIADIDAEFVATFDDQQLVGVSVPANVRTAGFVPMNVLLPTCAATVHHGGTGSWLTA
AIHGVPQIILSDADTEVHAKQLQDLGAGLSLPVAGMTAEHLRGAIERVLDPEPAYRLGA
ERM RDGMRTDPSPAQVVGICQDLAADRAARGRQPRRTAEPHLPR" (SEQ ID NO: 4)
gene 3462..4634
      /gene="megY"
CDS   3462..4634
      /gene="megY"
      /codon_start=1
      /transl_table=11

```

FIG. 7-3

11/70

```

/product="mycarose O-acyltransferase"
/translation="MVTSTNLDTTARPALNSLTGMRFAAFLVFFTHVLSRLIPNSYV
YADGLDAFWQTTGRVGVSEFFILSGFVLTWSARASDSVSWFRRRVCKLFPNHLVTAF
AAVVLFLVTGQAVSGEALIPNLLLIHAWFPALIEISFGINPVSWSLACEAFFYLCFPLF
LFWISGIRPERLWAAVAVFAAIWAVPVVADLLLPSSPPLIPGLEYSAIQDWFLYTFF
ATRSLEFILGIIILARILITGRWINVGLLPVLLFPVFFVASLFLPGVYAISSMMILP
LVLI IASGATADLQQKRTFMRNRVMVWLGDVSFALYMHFLVIVYGADLLGFSQTEDA
PLGLALFMIIPFLAVSLVLSWLLYRFVELPVMRNWARPASARRKPATEPEQTPSRR"
4651..5775 (SEQ ID NO: 5)
gene
/gene="megDII"
4651..5775
CDS
/gene="megDII"
/codon_start=1
/transl_table=11
/product="TDP-3-keto-6-deoxyhexose 3-aminotransaminase"
/translation="MTTYVWSYLLYERERADILDVQKVFASGSLILGQSVENFETE
YARYHGIAHCVGVDNGTNAVKLALLESVGVGRDDEVVTVSNTAAPTVAIDEIGARPVE
VDVRDEDYLMDTDLEAAVTPRTKAIVPVHLYGQCVDMTALRELADRGLKLVEDCAQ
AHGARRDGRLAGTMSDAAAFSFPYPTKVLGAYGDGAVVTNDDDETARALRLRYGMEE
VYYVTRTPGHNSRLDEVQAEILRRKLTRLDAYVAGRRVAQRYVDGLADLQDSHGLEL
PVVTDGNEHVFYVVRHPRRDEIIKRLRDGYDISLNISYPWPVHTMTGFAHLGVASG
SLPVTERLAGEIFSLPMYPSLPHDLQDRVIEAVREVITGL" (SEQ ID NO: 6)
5822..6595
gene

```

FIG. 7-4

12/70

```
CDS      /gene="megDIII"
5822..6595
/gene="megDIII"
/codon_start=1
/transl_table=11
/product="daunosaminyl-N,N-dimethyltransferase"
/translation="MPNSHSTTSSTDVAPYERADIYHDFYHGRGKGYRAEADALVEVA
RKHTPQAATLLDVACGTGSHLVELADSFREVVGVDLSAAMLATAARNDPGRELHQGDM
RDFSLDRRFDVVTCTMFSSGTGYLVDEAELDRAVANLAGHLAPGGTLVVEPWFFPETERP
GWVGADLVTSDDRISRMSHTVPAGLPDRITASRMTIHYTVGSPEAGIEHFTVHVMTL
FARAAEQAFQRAGLSCSYVGHDLESPGLEVGVAEPGR" (SEQ ID NO: 7)
6592..7197
gene      /gene="megDIV"
6592..7197
CDS      /gene="megDIV"
/codon_start=1
/transl_table=11
/product="TDP-4-keto-6-deoxyhexose 3,5-epimerase"
/translation="MRVEELGIEGVFTTTPQTFADERGVGTAYQEDVFVAALGRPLF
PVAQVSTTRSRRGVVRGVHFTTMPGSMKYVYCARGRAMDFAVDIRPGSPTFGRAEPV
ELSAESMVGLYLPVGMGHLFVSLEDDTTLVYLSAGYVPDKERAVHPLDPELALPIPA
DLDLVMSEDRVAPTTLREARDQGILPDYAACRAAAHRVVRT" (SEQ ID NO: 8)
7220..8206
gene
```

FIG. 7-5

13/70

CDS	<pre> /gene="megDV" 7220..8206 /gene="megDV" /codon_start=1 /transl_table=11 /product="TDP-4-keto-6-deoxyhexose 4-ketoreductase" /translation="MVVLGASGFLGSAVTHALADLPVRVRLVARREVVPSPGAVADYE THRVDLTEPGALAEVVADARAVFPFAAQIRGTSGWRISEDVVAERTNVGLVRDLIAV LSRSPHAPVVVFPGSNTQVGRVTAGRVIDGSEQDHPEGVYDRQKHTGEQLLKEATAAG AIRATSLRLPPVFGVPAAGTADDRGVVSTMIRRALTGQPLTMWHDGTVRRELLYVTDA ARAFVTALDHADALAGRHFLLGTGRSWPLGEVFQAVSRSVARHTGEDPVPVSVPPPA HMDPSDLRSVEVDPAFTAVTGWRATVTMAEAVDRTVAALAPRRAAAPSEPS" complement (8228..9220) </pre>	(SEQ ID NO: 9)
gene		
CDS	<pre> /gene="megDVII" complement (8228..9220) /gene="megDVII" /codon_start=1 /transl_table=11 /product="TDP-4-keto-6-deoxyhexose 2,3-reductase" /translation="MGTTGAGSARVRVGRSALHTSRLWLGTVNFSGRVTDDALRLMD HALERGVNCIDTADIYGWRLYKGHTEELVGRWFAQGGRRRETVLATKVGSEMSERVN DGGLSARHIVAACENSRLRRLGVDHIDIYQTHHIDRAAPWDEVWQAAEHLVSGKVGYY GSSNLAGWHIAAAQESAARRNLLGMISHQCLYNLAVRHPELDVLPAQAQYGVGVFAWS </pre>	

FIG. 7-6

14/70

```

gene
    PLHGGLLSGVLEKLAAGTAVKSAQGRAQVLLPAVRPLVEAYEDYCRRLGADPAEVGLA
    WVLSRPGILGAVIGPRTPEQLDSALRAAELTLGEEELRELEAIFPAPAVDGPVP"
    complement (9226..10479) (SEQ ID NO: 10)
    /gene="megBV"
CDS
    complement (9226..10479)
    /gene="megBV"
    /codon_start=1
    /transl_table=11
    /product="TDP-mycarose glycosyltransferase"
    /translation="MRVLLTSFAHRTHFQGLVPLAWALHTAGHDVVRVASQPELTDVVV
    GAGLTSVPLGSDHRLFDISPEAAQVHRYTTDLDFARRGPELRSWEFLHGIEEATSRF
    VFPVNNDSFVDELVEEFAMDWRPDLVLWEPFFAGAVAAKACGAHARLLWGSDLTGY
    FRSRSQDLRGQRPADDRPDLGGWLTEVAGRFGLDYSEDLAGQWSVDQLPESFRLET
    GLESVHTRTLPYNGSSVVPQWLRTSDGVRRVCFTGGYSALGITSNPQEEFLRTLATLAR
    FDGEIVVTRSGLDPASVPDNVRLVDFVPMNILLPGCAAVIHHGGAGSWATALHHGVPQ
    ISVAHEWDCVLRGQRTAELGAGVFLRPDEVDAATLWQALATVVEDRSHAENAEKLRQE
    ALAAPTPAEVVPVLEALAHQHRADR" (SEQ ID NO: 11)
    complement (10483..11424)
    /gene="megBIV"
CDS
    complement (10483..11424)
    /gene="megBIV"
    /codon_start=1
    /transl_table=11

```

FIG. 7-7

15/70

```
/product="TDP-4-keto-6-deoxyhexose 4-ketoreductase"  
/translation="MTRHVTLLGVSGFVGSALLREFTTHPLRLRAVARTGSRDQPPGS  
AGIEHLRVDDLLEPGRVAQVADTDVVHLVAYAGGSTWRSAAATVPEAERVNAGIMRD  
LVAALRARPAPVLLFASTTQAANPAAPSRYAQHKIEAERILLRQATEDGVVDGVILR  
LPAIYGHSGPSGQTGRGVVTAMIRRALAGEPITMWHEGSVRRNLLHVEDVATAFTAAAL  
HNHEALVGDVWTPPSADEARPLGEIFETVAASVARQTGNPAVPVSVPPPENAEANDFR  
SDDFDSTEFRTL TGWHPRVPLAEGIDRTVAALISTKE" (SEQ ID NO: 12)  
  
gene  
12181..22821  
/gene="megAI"  
  
CDS  
12181..22821  
/gene="megAI"  
/note="polyketide synthase"  
/codon_start=1  
/transl_table=11  
/product="megalomycin 6-deoxyerythronolide B synthase 1"  
/translation="MVDVPDLLGTRTPHPGPLFPWPPLCGHNEPELRARARQLHAYLE  
GISEDVVAVGAALARETRAQDGP HRAVVVASSVTELTAAALAAQGRPHPSVVRGVA  
RPTAPVVFVLPGQGAQWP GMATRL LAESPFAAAMRACERAFDEVTDWSLTVELDSPE  
HLRRVEVVQPALFAVQTSIALALWRSEFGRPD AVLGH SIGELAAAEVCGAVDVEAAARA  
AALWSREMVPLVGRGDMAAVALSPAELAAARVERWDDDDVVPAGVNGPRSVLLTGAPPEPI  
ARRVAELAAQGVRAQVVNVSMAAHSAQVDAVAEGMRSALTWFAPGSDSDVPYYAGLTGG  
RLDTRELGADHWPRSFRLPVRFDEATRAVLELQPGTFIESSPHPVLAASLQQTLDEVG  
SPA AIVPTLQRDQGLRRRFLLA VAQA YTG GVTVDWTAA YPGVTPGHLPSA VAVETDEG
```

FIG. 7-8



16/70

PSTFDDWAAPDHVLRARLLEIVGAETAALAGREVDARATFRELGLDSVLAVQLRTRLA  
TATGRDLHIAMLYDHPTPHALTEALLRGPQEEPGRGEETAHPTAEAPDEPVAVVAMAC  
RLPGGVTSPEEFWELLAEGRDAVGGLPTDRGWDLDLSLPHDPTRSGTAHQAGGFLTG  
ATSFDAAFFGLSPREALAVEPQQORITLELSWEVLERAGIPPTSLRTSRTGVFVGLIPQ  
EYGPRLAEGGEGVEGYLMTGTTTSVASGRVAYTLGLEGPAISVDTACSSSLVAVHLAC  
QSLRRGESTMALAGGVTVMPPTPGMLVDFSRMNSLAPDGRSKAFSAADGFGMAEGAGM  
LLLERLSDARRHGHPVLAVIRGTAVNSDGASNGLSAPNGRAQVRVIRQALAESGLTPH  
TVDVVEHGTGTRLGDPTEARALSDAYGGDREHPLRIGSVKSNIGHTQAAAGVAGLIK  
LVLAMQAGVLPRTLHADEPSPEIDWSSGAISLLQEPAAWPAGERPRRAGVSSFGISGT  
NAHAIIEEAPPTGDDTRPDRMGVPVWPVLSASTGEALRARAARLAGHLREHPDQDLDLDD  
VAYSLATGRAALAYRSGFVPADASTALRILDELAAGSGDAVTGTARAPQVRVVFVFPFG  
QGWQWAGMAVDLLDGDVPFASVLRCADALEPYLDFFIVPFLRAEAQRRTPDHTLSTD  
RVDVVQPVLFAVMVSLAARWRAYGVEPAAVIGHSQGEIAAACVAGALSLLDDAARAVAL  
RSRVIAATMPGNGAMASIAASVDEVAARIDGRVEIAAVNGPRAVVVSGDRDDLDRLVAS  
CTVEGVRAKRLPVDYASHSSHVEAVRDALHAELGEFRPLPGFVPFYSTVTGRWVEPAE  
LDAGYWFRNLHRVRFADAVRSLADQGYTTFLEVSHPVLTAEIIGEDRGGDLVAV  
HSLRRGAGGPVDFGSALARAFVAGVAVDWESAYQGAGARRVPLPTYPFQREFWLEPN  
PARRVADSDDVSSLRYRIEWHPTDPGEPGRLDGTWLLATYPGRADDRVEAARQALESA  
GARVEDLVVEPRTGRVDLVRRLDAVGPVAGVLCLEFAVAEPAAEHSPPLAVTSLSDTLDL  
TQAVAGSGRECPIWVVTENAVAVGPFERLRDPAHGALWALGRVVALENPAVWGGLVDV  
PSGSVAELSRHLGTTLSGAGEDQVALRPDGTYARRWCAGAGGTGRWQPRGTVLVTGG  
TGGVGRHVARWLARQGTPLVLSRRGPDADGVVEELLTELADLGTATVTACDVTDRE  
QLRALLATVDDEHPLSAVFHVAATLDDGTVETLTGDRIERANRAKVLGARNLHELTRD

FIG. 7-9

17/70

ADLDAFVLFSSSTAAGAPGLGGYVPGNAYLDGLAQRRSEGLPATSVAWGTWAGSGM  
AEGPVADRFRRHGVMEMHPDQAVEGLRVALVQGEVAPIVVDIRWDRFLLAYTAQRPTR  
LFDTLDEARRAAPGPDAGPGVAALAGLPVGEREKAVLDLVRTHAAAVLGHASAEQVPV  
DRAFAELGVDSL SALELRNRLTTATGVRLATTTVFDPDVRTLAGHLAAELGGSGRE  
RPGGEAPTVAPTDEPIAIVGMACRLPGGVDSPEQLWELIVSGRDTASAAPGDRSWDPA  
ELMVSDTTGTRTAFGNFMPGAGEFDAFFGISPREALAMDPOQRHALETTWAEALENAG  
IRPESLRGTDTGTFVGM SHQGYATGRPKPEDEV DGYLLTGNTASVASGRIAYVLGLEG  
PAITVDTACSSSLVALHVAAGSLRSGDCGLAVAGGVSMAGPEVFEFSRQGALAPDG  
RCKPFSDEADGFGLEGESAFVVLQRLSVAVREGRRVLGVVGSVAVNQDGASNGLAAPS  
GVAQQRVIRRAWGRAGVSGGDVGVEAHGTGTRLGDPVELGALLGTYGVRGGVGPVV  
VGSVKANVGHVQAAAGVVGVIKVVLGLGRGLVGPVMVCRGGLSGLVDWSSGGLVVADGV  
RGWPGVDGVRRGVSAGFVSGTNAHVVAEAPGSVGAERPVEGSSRGLVGVVGGVV  
PVVLSAKTETALHAQARRLADHLETHPDVPMTDVVWTLTQARQFRDRRAVLLAADRTQ  
AVERLRGLAGGEPGTGVVSGVASGGGVVFVFPQGQGVGMARGLLSVPVFVESVVEC  
DAVVS SVVGFESVLGVLEGRSGAPSLDRVDVVPVLFVVMVSLARLWRWCVVPAAVVG  
HSQGEIAAAVVAGVLSVGDGARVVALARALRALAGHGGMASVRRGRDDVQKLLDSGP  
WTGKLEIAAVNGPDAVVSGDPRAVTELVEHCDGIGVRARTIPVDYASHSAQVESLRE  
ELLSVLAGIEGRPATVPFYSTLTGGFVDGTELDADYWYRNLRHPVRFHAAVEALAARD  
LTTFVEVSPHPVLSMAVGETLADVESA VTVGTLERD TDDVERFLTSLAEAHVHGVPVD  
WAAVLGSGTLVDLPTYPFQGRRFWLHPDRGPRDDVADWFHRVDWTATATDGSARLDGR  
WLVVVEGYTDDGWVVEVRAALAAAGGAEPVTTVEEVTD RVGDSDAVVSMLGLADDGA  
AETLALLRRLLDAQASTTPLWVVTVGAVAPAGPVQRPEQATVWGLALVASLERGHRWTG  
LLDL PQTPDPQLRPRLVEALAGAEDQAVRADAVHARRI VPTPVTGAGPYTAPGGTIL

FIG. 7-10

18/70

VTGGTAGLGAVTARWLAERGAEHLALVSRRGPGTAGVDEVVRDLTGLGVRVSVHSCDV  
 GDRESVGALVQELTAAGDVVRGVVHAAGLPQQVPLTMDPADLADVAVKVDGAVHLA  
 DLCPEAELELLFSSGAGVWGSARQGAAYAGNAELDAFARHRRDRGLPATSVAWGLWAA  
 GMTGDQEAVSFLRERGVPRMSVPRALSALEVLTAGETAVVADVDDWAAFAESYTS  
 RPRPLLHRLVTPAAAVGERDEPREQTLRDRLAALPRAERSAELVRLVRRDAAAVLGSD  
 AKAVPATTPFKDLGFDSLAAVFRNRRLAHTGLRLPATLVFEHPNAAAVADLLHDLRLG  
 EAGEPTPVRSVGAGLALEQALPDASDTERVELVERLERMLAGLRPEAGAGADAPTAG  
 DDLGEAGVDELLDALEREELDAR" (SEQ ID NO: 13)

misc_feature	12505..13470	/gene="megAI"
		/function="AT-L"
misc_feature	13576..13791	/gene="megAI"
		/function="ACP-L"
misc_feature	13849..15126	/gene="megAI"
		/function="KS1"
misc_feature	15427..16476	/gene="megAI"
		/function="AT1"
misc_feature	17155..17694	/gene="megAI"
		/function="KR1"

FIG. 7-11

19/70

```
misc_feature 17947..18207
              /gene="megAI"
              /function="ACP1"
misc_feature 18268..19548
              /gene="megAI"
              /function="KS2"
misc_feature 19876..20910
              /gene="megAI"
              /function="AT2"
misc_feature 21517..22053
              /gene="megAI"
              /function="KR2"
misc_feature 22318..22575
              /gene="megAI"
              /function="ACP2"
gene          22867..33555
              /gene="megAII"
CDS           22867..33555
              /gene="megAII"
              /note="polyketide synthase"
              /codon_start=1
              /transl_table=11
              /product="megalomycin 6-deoxyerythronolide B synthase 2"
```

FIG. 7-12

20/70

/translation="MTDNDKVAEYLRRTATLDLRAARKRLRELQSDPIAVVGMACRLPG  
GVHLPQHLWDLRLRQGHETVSTFPTGRGWDLAGLFHPDPPHGTSYVDRGGFLDDVAGE  
DAEFFGISPREATAMPQORLLLETSWELVESAGIDPHSLRGTPGTGFLGVARLGYGE  
NGTEAGDAEGYSVTGVAPAVASGRISYALGLEGPSISVDTACSSSLVALHLAVESLRL  
GESSLAVVGGAAVMATPGVFVDFSRQALAAADGRSKAFGAAADGFGFSEGVSLVLLER  
LSEAESNGHEVLAVIRGSALNQDGASNGLAAPNGTAQRKVIRQALRNCGLTPADVDAV  
EAHGTGTTLGDP I EANALLDTYGRDRDPDHP L W L G S V K S N I G H T Q A A A G V T G L L K M V L  
ALRHEELPATLHVDEPTPHVDWSSGAVRLATRGRPWRRGRDRPRRAGVSAFGISGTNAH  
VIVEEAPERTTERTVGGDVGPVPLVVSARSAALRAQAAQVAELVEGSDVGLAEVGRS  
LAVTRARHEHRAAVVASTRAEAVRGLREVAAVEPRGEDTGTVAETSGRTVVFLFPGQ  
GSQWVGMAELLDSAPAFADTIRACDEAMAPLQDWSVSDVLRQEPGAPGLDRVDVVQP  
VLFVVMVSLARLWQSYGVTPAAVVGHQSQGEIAAAHVAGALSADAAARLVVGRSRLRS  
LSGGGMSAVALGEAEVRRRLRSWEDRISVAAVNGPRS VVAGEPEALREWG REREAE  
GVRVREIDVDYASHSPQIDRV RDEL L T V T G E I E P R S A E I T F Y S T V D V R A V D G T D L D A G  
YWYRNLR E T V R F A D A M T R L A D S G Y D A F V E S P H P V V S A V A E A V E E A G V E D A V V V G T L  
SRGDGGPGAFLRSAATAHCAGVDVDWTPALPGAATIPLTYPFQRPYWLRSAPAPA  
SHDLAYRVSWTPI TPPGDGVL DGDWL VVHPGGSTGWVDGLAAAITAGGGRVVAHPVDS  
VTSRTGLAEALARRDGTFRGVLSWVATDERHVEAGAVALLTLAQALGDAGIDAPLWCL  
TQEAVRTPPVDGLARPAQAALHGFAQVARLELARRFGGVLDPATVDAAGTRLVAAVL  
AGGGEDVVAVRGDRLYGRRLVRATLPPPPGGFTPHGTVLVTGAAGPVGGRLARWLAER  
GATRLVLPGAHPGEEELLTAIRAAGATAVCEPEAEALRTAIGGELPTALVHAETLTNF  
AGVADADPEDFAATVAAKTALPTVLAEVLGDHRLEREVYCSSVAGVWGVGMAYAAAG  
SAYLDALVEHRRARGHASASVAWTPWALPGA VDDGR L R E R G L R S L D V A D A L G T W E R L L

FIG. 7-13

21/70

RAGAVSAVADVDSVFTTEGFAAIRPTPLFDELLDRRGDPDGAPVDRPGEPAGEWGRR  
IAALSPQEORETLLTLVGETVAEVLGHETGTEINTRRAFSELGLDSLGMALRQRLAA  
RTGLRMPASLVFDHPTVTALARYLRLVVGSDSDPTPVRVFGPTDEAEPVAVVGICRF  
PGGIATPEDLWRVVSEGTSITTGFPTRDGRWDLRRLYHPDDHPGTSYVDRGGFLDGAP  
DFDPGFFGITPREALAMPQQORLTLEIAWEAVERAGIDPETLLGSDTGTVFGMNGQSY  
LQLLTGEGDRLNGYQGLNSASVLSGRVAYTFGWEGPALTVDTACSSSLVAIHLAMQS  
LRRGECSLALAGGVTVMADPYTFVDFSAQRGLAADGRCKAFSAQASGFALAEGVAAALV  
LEPLSKARRNGHQVLAVLRGSVAVNQDGASNGLAAPNGPSQERVIRQALTASGLRPADV  
DMVEAHGTGTELGDPIEAGALIAAYGRDRDRPLWLGSVKTNIGHTQAAAAGAGVIKAV  
LAMRHGVLPRSLHADELSPHIDWADGKVEVLREARQWPGERPRRAGVSSFGVSGTNA  
HVI VEEAPAEPPPEPVAAPGGPLPFVLHGRSVQTVRSQARTLAEHLRTTGHRDLADT  
ARTLATGRARFDVRAAVLGTDRREGVCAALDALAQDRSPDVVAPAVFAARTPVLVFPG  
QGSQWGMARDLLDSSEVFAESMGRCAEALSPYTDWDLDDVVVRGVGDPDPYDRVDVLQ  
PVLFAVMVSLARLWQSYGVTPGAVVGHSGQGEIAAAHVAGALSLADAARVVVALRSRVLR  
ELDDQGMVSVGTSRAELDSVLRWRWDGRVAVAAVNGPGTLVVAGPTAELDEFLLAVAEA  
REMRPRRIAVRYASHSPEVARVEQRLAAELGTVTAVGGTVPLYSTATGDLDDTTAMDA  
GYWYRNLRQPVLFEEHAVRSLLERGFETFIEVSPHPVLLMAVEETAEDAERPVTGVPTL  
RRDHDGPSEFLRNLLGAHVHGVVDVLRPAVAHGRVLDLPTYPFDRQRLWPKPHRRADT  
SSLGVRDSTHPLLHAAVDVPGHGGAVFTGRLLSPDEQQWLTQHVVVGRNLPVGSVLVDL  
ALTAGADVGPVLEELVLQQLVLTAAAGALLRLSVGAADDEDGRRPVEIHAAEDVSDPA  
EARWSAYATGTLAVGVAGGGRDGTQWPPPGATALTLDHYDTLAEELGYEYGPFAFQALR

FIG. 7-14



22/70

AAWQHGDVVYAEVSLDAVEEGYAFDPVLLDAVAQTFGLTSRAPGKLPFAWRGVTLHAT  
GATAVRVATPAGPDAVALRVTDPTGQLVATVDALVVRDAGADRDPGRDGDHLHRL  
WVRLATPDPTPAAVVHVAADGLDDLLRAGGPAPQAVVRYRPDGGDDPTAEARHGVLWA  
ATLVRRWLDDDRWPATTLVVATSAGVEVSPGDDVPRPGAAA VGVLRCAQAESPDRE  
LVDGDPETPPAVPDNPQLAVRDGAVFVPRLTPLAGVPVAVADRAYRLVPGNGGSI  
EAVAFAPVPDADRPLAPEEVRVAVRATGVNFRDVL LALGMYPEPAEMGTEASGVVTE  
VSGVRRFTPGQAVTGLFQGAFGPVAVADHRLLTVPDGWRAVDAAA VPIAFTTAHYAL  
HDLAGLQAGQSVLVHAAAGGVGMAAVALARRAGAEVFATASPAKHPTLRALGLDDH  
IASSRESGGERPAARTGGRGVDVVLNSLTGDLDESARLLADGGVFVEMGKTDLRPA  
EQFRGRYPFDLAEAGPDRLGEILLEEVGLLAAGALDRLPVSVWELSAAPALTHMSR  
GRHV GKLVLTPQAPVHPDGTVLVTGGTGLGRLVARHLVTGHGVPHLLVASRRGP  
AAPGAAELRADVEGLGATIEIVACDTADREALAALLDSIPADRPLTGVVHTAGVLAD  
GLVTSIDGTATDQVLRAKVDAAWHLHDLTRDADLSFFVLFSSAASVLAGPGQGVYAA  
ANGVLNALAGQRRALGLPAKALGWLWAQASEMTSGLGDRIARTGVAALPTERALALF  
DAALRSGEVLFPLSVDRSALRRAEYVPEVLRGAVRSTPPRAANRAETPGRGLLDRL  
VGA PETDQVAA LAELVRSHAAAVAGYDSADQLPERKAFKDLGFDLSLA AVELRNRL  
GVTGVRLPSTLVFDHPTPLAVAEHLRSELFAD SAPDVGVGARLDDLERALDLPDAQGH  
ADV GARLEALLRRWQSRPPETEPVTISDDASDDELFSMLDRRLGGGDV" (SEQ ID NO: 14)

misc\_feature 22957..24237  
/gene="megAII"  
/function="KS3"  
misc\_feature 24544..25581  
/gene="megAII"  
/function="AT3"

FIG. 7-15

23/70

misc_feature	26230..26733 /gene="megAII" /function="KR3" (inactive)
misc_feature	26998..27258 /gene="megAII" /function="ACP3"
misc_feature	27393..28590 /gene="megAII" /function="KS4"
misc_feature	28897..29931 /gene="megAII" /function="AT4"
misc_feature	29953..30477 /gene="megAII" /function="DH4"
misc_feature	31396..32244 /gene="megAII" /function="ER4"
misc_feature	32257..32799 /gene="megAII" /function="KR4"
misc_feature	33052..33312 /gene="megAII" /function="ACP4"

FIG. 7-16

24/70

gene  
33666..43271  
/gene="megAIII"  
CDS  
33666..43271  
/gene="megAIII"  
/note="polyketide synthase"  
/codon\_start=1  
/transl\_table=11  
/product="megalomicin 6-deoxyerythronolide B synthase 3"  
/translation="MSESSGMTEDRLRRYLKRTVAELDSVTGRLDEVEYRAREPIAVV  
GMACRFPGGVDSPAEAFWEFIRDGGDAIAEAPTDRGWPPAPRRLGGLLAEPGAFDAAF  
FGISPREALATDPQQRLMLEISWEALERAGFDPSLRGSAGGVFTGVGAVDYGPRPDE  
APEEVLGYVGIGTASSVASGRVAYTLGLEGPVTVDTACSSGLTAVHLAMESLRRDEC  
TLVLGGVTVMSSPGAFTFRSQGLAEDGRCKPFSRAADGFGLAEGAGVLVLQRLSV  
ARAEGRPVLAFLRGSAINQDGASNGLTAPSGPAQRRVIRQALERARLRPVDVDYVEAH  
GTGTRLGDPIEAHALLDTYGADREPGRPLWVGSVKSNIQHTQAAAGVAGVMKTVLALR  
HREIPATLHFDEPSPHVDWDRGAVSVSETRPWPVGERPRRAGVSSFGISGTNAHVIV  
EEAPSPQAADLDPTPGPATGATPGTDAAPTAEPGAEEAVALVFSARDERALARQAARLA  
DRLTDDPAPSLRDTAFTLVTRRATWEHRAVVVGGGEEVLAGLRAVAGGRPVDGAVSGR  
ARAGRVRVLVFPGQGAQWQGMARDLLRQSPTFAESIDACERALAPHVDWSLREVL DGE  
QSLDPVDVVPVLFVVMVSLARLWQSYGVTPGAVVGHSHSQGEIAAAHVAGALS LADAAR  
VVALRSRVLRLGGHGMASFGLHPDQAAERIAREFAGALTVASVNGPRS VVLAGENGP  
LDELIAECEAEGVTARRIPVDYASHSPQVESLREELLAALAGVRPVSAGIPLYSTLTG  
QVIETATMDADYWFANLREPVRFDATRQLAEAGFD AFVEVSPHPVLT VGV EATLEAV  
LPPDADPCVTGTLRRERGGLAQFHTALAEAYTRGVEVDWRTAVGEGRPVDLPVYPFQR

FIG. 7-17

25/70

QNEWLPVPLGRVPDGTGDEWRYQLAWHPVDLGRSSLAGRVLVVTGAAVPPAWTDVVVRDG  
LEQRGATVVLCTAQSRARIGAAALDAVDGTALSTVVSLLALAEAGAVDDPSLDTLALVQ  
ALGAAGIDVPLWLVTTRDAAAVTVGDDVDPAQAMVGGLGRVVGVESPARWGGLVDLREA  
DADSARSLAAIILADPRGEEQFAIRPDGVTVARLVPAPARAAGTRWTFRGTVLVTGGTG  
GIGAHLARWLAGAGAEHLVLLNRRGAEEAAGAADLRDELVALGTGVTITACDVADRDL  
AAVLDAARAQGRVVTAVFHAAGISRSSTAVQELTESEFTEITDAKVRGTANLAELCP  
DALVLFSSNAAVWGSPGLASYAAGNAFLDAFARRGRSGLPVTSLAWGLWAGQNMA  
EGGDYLRSQLRAMDPQRAIEELRTTLDAGDPWVSVDLDRERFVELFTAARRRPLFD  
ELGGVRAGAEETQESDLARRLASMPEAERHEHVARLVRAEVAAVLGHGTPTVIERDV  
AFRDLGFDSTAVDLNRNLAAVTGVRVATTIVFDHPTVDRLTAHYLERLVEPEATTP  
AAAVVPQAPGEADEPIAIVGMACRLAGGVRTPDQLWDFIVADGDAVTEMPSDRSWDLD  
ALFDPDPERHGTSYSRHGAFLDGAADFDAFFGISPREALAMDPOQRQVLETTWELFE  
NAGIDPHSLRGDTGVFLGAAQQGYGQNAQVPKESEGYYLLTGSSAVASGRIAYVLGL  
EGPAITVDTACSSSLVALHVAAGSLRSGDCGLAVAGGVSVMAGPEVTFEFSRQGALAP  
DGRCKPFSDAQDGEFAEGVAVVLLQRLSVAVREGRRVLGVVVGSAVNQDGASNGLAA  
PSGVAQQQVIRRAWGRAGVSGGDVGVVEAHGTGTRLGDPVELGALLGTYGVGRGGVGP  
VVVGSVKANVGHVQAAAGVVGVVKVVLGLGRGLVGPMVCRGGLSGLVDWSSGGLVVAD  
GVRGWPGVDGVRRGVS AFGVSGTNAHVVAEAPGSVGAERPVEGSSRGLVGVAGG  
VVPVVL SAKTETALTELARRLHDAVDDTVALPAVAATLATGRAHLPYRAALLARDHDE  
LRDRLRAFTTGSAA PGVVS GVASGGVVFVFPQGQWVGMARGLLSVPVFVESVVEC  
DAVSSVVGFSVLGVLEGRSGAPSLDRVDVQPVLFVVMVSLARLWRWCGVVPAAVVG  
HSQGEIAAAVVAGVLSVGDGARVVVALRARALRALAGHGMVSLAVSAERARELIAPWS  
DRISVAAVNSPTSVVVS GDDPQALAAALVAHCAETGERAKTLPVDYASHSAHVEQIRDTI  
LTDLADV TARRPDVALYSTLHGARGAGTMDARYWYDNLRSVPVRFDEAVEAAVADGYR

FIG. 7-18

26/70

VFVEMSPHPVLTAAVQEIDDETVAIGSLHRDTGERHLVAELARAHVHGVPVDWRAILP  
 ATHPVLPNYPFEATRYWLAPTAADQVADHRYRVDWRPLATTPAELSGSYLVFGDAPE  
 TLGHSVEKAGGLVPVAPDRESLAVALDEAAGRLAGVLSFAADTATHLARHRLLEGEA  
 DVEAPLWLVTSGGVALDDHDPIDCDQAMVWIGIGRVMGLETPHRWGGLVDTVVEPTAED  
 GVVFAALLAADDEDQVALRDGIRHGRRLVRAPLTTNRNARTPAGTALVTGGTGALGG  
 HVARYLARSQVTDLVLLSRSGPDAPGAAELAAELADLGAEPVEACDVTGPRRLRALV  
 QELREQDRPVRIVVHTAGVPDSRPLDRIDELESVSAKVTGARLLDELCPDADTFVLF  
 SSGAGVWGSANLGAYAAANAYLDALHRRRQAGRAATSVAWGAWAGDMATGDLDTGLT  
 RRGLRAMAPDRALRACTRRWTTHTDTCVSVADVVDWDRFAVGFTAAARPRPLIDELVTSAP  
 VAAPTAAAPVPAMTADQLLQFTRSHVAAIILGHQDPPDAVGLDQPFTELGFDSLTAAGL  
 RNQLQQATGRTLPAALVFQHTVRRRLADHLLAQQLDVGTAPVEATGSVLRDGYRRAGQT  
 GDVRSYDLLANLSEFRERFTDAASLGGQLELVLDLADSGSPVTVICCAGTAALSGPHE  
 FARLASALRGTPVRALAQPGYEAGEVPASMEAVLGQADAVLAAQGDTPFVLVGHHS  
 AGALMAYALATELADRGHPPRGVVLLDVYPPGHQEAHVHAWLGELTAALFDHETVRMDD  
 TRLTALGAYDRLTGRWRPRDTGLPTLVVAASEPMGEWPDGQSTWPFQHDRVTVPGD  
 HFMSMVQEHADAIA RHIDAWLSGERA" (SEQ ID NO: 15)

```
misc_feature 33780..35027
              /gene="megAIII"
              /function="KS5"
misc_feature 35385..36419
              /gene="megAIII"
              /function="AT5"
```

FIG. 7-19

27/70

misc\_feature 37068..37604  
/gene="megAIII"  
/function="KR5"  
misc\_feature 37860..38120  
/gene="megAIII"  
/function="ACP5"  
misc\_feature 38187..39470  
/gene="megAIII"  
/function="KS6"  
misc\_feature 39795..40811  
/gene="megAIII"  
/function="AT6"  
misc\_feature 41406..41936  
/gene="megAIII"  
/function="KR6"  
misc\_feature 42168..42425  
/gene="megAIII"  
/function="ACP6"  
misc\_feature 42585..43271  
/gene="megAIII"  
/function="TE"  
gene 43268..44344  
/gene="megCII"

FIG. 7-20



28/70

CDS

43268..44344  
 /gene="megCII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TDP-4-keto-6-deoxyglucose 3,4-isomerase"  
 /translation="MNNTDRAVLGRRLLQIRGLYWGYSNGDPYPMLLCGHDDDDPHRW  
 YRGLGSGVRRSRTETWVTDHATAVRVLDLDDPTFTTRATGRTPEWMRAAGAPASTWAQP  
 FRDVHAASWDAELPDPQVEEDRLTGLLPAPGTRLDLVRDLAWPMASRGVGADDDPDVLR  
 AAWDARVGLDAQLTPQPLAVTEAAIAAVPGDPHRRALFTAVEMTATAFVDAVLAVTAT  
 AGAAQRLADDDVAARLVAEVLRLHPTAHLERRTAGTETVGEHTVAAGDEVVVVAA  
 ANRDAGVFADPDRLDPDRADADRALSAQRGHPGRLEELVVVLTAAALRSVAKALPGLT  
 AGGPVVRRRRSPVLRATAHCPVEL" (SEQ ID NO: 16)

gene

44355..45623  
 /gene="megCIII"  
 44355..45623  
 /gene="megCIII"  
 /codon\_start=1  
 /transl\_table=11  
 /product="TDP-desosamine glycosyltransferase"  
 /translation="MRVVFSSMASKSHLFGVPLAWAFRAAGHEVRVVASPALTDDIT  
 AAGLTAVPVGTDVDLVDFMTHAGYDIIIDYVRSLSERDPATSTWDHLLGMQVLTPT  
 FYALMSPDSLVEGMISFCRSWRPDWSSGPQTFAASIAATVTGVAHARLLWGPDITVRA  
 RQKFLGLLPGQPAAHREDPLAEWLTSVERFGGRVPQDVEELVVGQWTIDPAPVGMRL

CDS

FIG. 7-21

29/70

DTGLRTVGMRYVDYNGPSVVPDWLHDEPTRRRVCLTLGISSRENSIGQVSVDDLLGAL  
GDVDAEIIATVDEQQLEGVAHV PANIRTVGFVPMHALLPTCAATVHHGGPGSWHTAAI  
HGVPQVILPDGWDTGVR AQRTE DQAGIALPVPELTSDQLREAVRRVLDDPAFTAGAA  
RMRADMLAEPSPA EVVDVCAGLVGERTAVG" (SEQ ID NO: 17)

gene

45620..46591

/gene="megBII"

CDS

45620..46591

/gene="megBII"

/codon\_start=1

/transl\_table=11

/product="TDP-4-keto-6-deoxyglucose 2,3 dehydratase"

/translation="MSTDA THVRLGRCALLTSRLWLGTAA LAGQDDADAVRLLDHARS  
RGVNCLDTADDD SASTSAQVAEESVGRWLAGDTGRREETVLSVTGVPPGGQVGGGL  
SARQIIASCEGSLRRRLGVDHVDVLHLPRVDRVEPWDEVWQAVDALVAAGKVCYVGSSG  
FPGWHIVAAQEHAVRRHRLGLVSHQC RYDLTSRHPELEVLPA AQAYGLGVFARPTRLG  
GLLGGDGP GAAARASGQPTALRS AVEAYEVEFCRDLGEHPAEVALAWVLSRPGVAGAV  
VGARTPGR LDSALRACGVALGATELTALDGI FPGVAAAGAAPEAWLR"

(SEQ ID NO: 18)

complement (46660..47403)

/gene="megH"

complement (46660..47403)

/gene="megH"

/note="putative thioesterase"

gene

CDS

FIG. 7-22

30/70

```

/codon_start=1
/transl_table=11
/product="TEII"
/translation="MNTWLRFGSADGHRARLYCFPHAGAAADSYLDLARALAPEVDV
WAVQYPGRQRRDERALGTAGEIADEVAAVLRDLVGEVPFALFGHSMGALVAYETARR
LEARPGVRPLRLFVSGQTAPRVHERRTDLPDEDGLVEQMRRLGVSEAAALADQGLDMS
LPVLRADHRVLRSYAWQAGPPLRAGITTLCGDTPLTTFVEDAQRWLPYSVVPGRTRTF
PGGHFYLADHVGEVAESVAPDLLRLTPTG" (SEQ ID NO: 19)
gene      complement (47411..>47981)
          /gene="megF"
CDS       complement (47411..>47980)
          /gene="megF"
          /codon_start=1
          /transl_table=11
          /product="C-6 hydroxylase"
          /translation="IRVQDDADRLSRDELTSIALVLLAGFEASVSLIGTYLLLT
          HPDQLALVRKDPALLPGAVEEILRYQAPPETTRFATAEVEIGGVTIPAYSTVLIANG
          AANRDPGQFPDPDRFDVTRDSRGHLTFGHGIHYCMGRPLAKLEGEVALGALFDRFPKL
          SLGFPSDEVVWRRSLLLRGIDHLPVRPNG" (SEQ ID NO: 20)
BASE COUNT      5962 a 16875 c 18045 g 7099 t
ORIGIN
      1 ctcgagccga tgctcggcgg cgcggtgggc caaccagtcg tggacgtcgt cggaggcggt
     61 gggagggtccg ccgtgccgag tcaggaaacg tattgccgat tgtgtggatt ccggagtcgc

```

FIG. 7-23

31/70

121 atgaccgttg accgatccc ccatagcct ctccgtgat gtcgtgggcg gtccgtgcgg  
181 taccgcccgg actgacattc gtcgatcaag accccgccca gtgtagggt gtgtagggct ccccccgcga  
241 cgggagaagg tccgtcgaac aacttccggg tgaccgggtcg ccggcgtcgg tgaaacgggc  
301 gtcggagcac ccgatcattg ctgtcgggtga acttcctaac tgtcggcgcg cacatcttct  
361 tgaccgggtg gtcccggtgt atgacgcgtt cccggcccgt ctggaactgt gcgtgggact  
421 gaccgggtgc ggcgtgtttt ggcgtgtttt cgaactgcgg attcgtcgat cgcgcagggtg  
481 ggagcgggtg gctgaccggg atgatctgca atcatggcg tcaatgacga tctcttgtag  
541 catggtccgc gccgaggtc\* cgacaggccc gaaacgccc gcatccagcc tggtcgacga  
601 cgtcgacatc accgtgcaag ccgcgatgac accgacacca cgccatgctg gtgcccact  
661 ggaagggtg cgcgatcagg gaaatggccg tgtcactaga cagacgcaa acagctgtcc  
721 gggcctgcgg aaacagcatc gatctgcgtc agccgttcat agcccggcg tgcccgcctt  
781 ggaatccgt gccaccggtc gtccgcagtg acgatcgcgg acccgggttt cgagacagca  
841 ggtagtaggc gatgcaggcg tttcgtctcg cgcgggacgc gtccgactag gtggaatccg  
901 tcacagtctt caatccggga gcgttctatg gcagtggcg atcgaaggcg gctgggcccgg  
961 gaggtagcaga tggcccgggg tcttactgg gggttcggtg ccaacggcga tctgtactcg  
1021 atgctcctgt ccggacggga cgacgacccc tggacctggt acgaacggtt gcgggcccgc  
1081 ggacggggac cgtacgccag tcgggcccga acgtgggtgg acgttgacca ccggaccgcc  
1141 gccgaggtgc tcgccgatcc gggcttcacc caccggcccgc cccgacgctgc ccggtggatg  
1201 caggtagggc actgcccgcc gccctcctgg gccggcccct tccgggagtt ctacgcccgc  
1261 accgaggacg cggcgtcggg gacagtggac gccgactggc tccagcagcg gtgcgccagg  
1321 ctggtgaccg agctggggtc gcgcttcgat ctcgtgaacg acttcgcccg ggagggtcccg  
1381 gtgctggcgc tcggtaccgc gccgcactc aagggcgtgg accccgaccg tctccgggtcc

FIG. 7-24

32/70

1441 tggacctcgg cgaccggggt atgcctggac gcccagggtca gcccgcaaca gctcgcggtg  
1501 accgaacagg cgctgaccgc cctcgacgag atcgacgcgg tcaccggcgg tcgggacgcc  
1561 gcggtgctgg tgggggtggt ggcggagctg gcggccaaca cggtgggcaa cgccgtcctg  
1621 gccgtcacgg agcttcccga actggcggca cgactgccc acgaccgga gaccgcgacc  
1681 cgtgtggtga cggagggtgc cgggacgagt cccggcgtcc acctggaacg ccgacccgcc  
1741 gcgtcggacc gccgggtggg cggggtcgac gtcccgaccg gtggcgaggt gacagtggtc  
1801 gtcgcccgcg cgaaccgtga tcccagggtc ttcaccgata ccgaccggtt cgacgtggac  
1861 cgtggcgggc acgccgagat cctgtcgtcc cggcccggct cgccccgcac cgacctcgac  
1921 gccctggtgg ccaccctggc cacggcgggc ctgcggggccg ccgcgccgggt gttgccccgg  
1981 ctgtcccgtt ccgggcccgt gatcagacga cgtcgggtcac ccgtcgcccg tggctcagc  
2041 cgttgcccgg tcgagctgta gaggaagaac gatgcgcgtc gtgttttcat cgatgggtgt  
2101 caacagccat ctgttcgggc tggccccgt cgcaagcgcc ttccaggcgg ttccgacaga  
2161 ggtaagggtc gtcgcctcgc cggccctgac cgacgacgtc accggtgccg gtctgaccgc  
2221 cgtgcccgtc ggtgacgacg tggaaacttgt ggagtggcac gccacgcgg gccaggacat  
2281 cgtcgagtac atgcggacc tcgactgggt cgaccagagc cacaccacca tgtcctggga  
2341 cgacctcctg ggcatgcaga ccaccttcac ccgaccttc ttcgccctga tgagccccga  
2401 ctcgctcatc gacgggatgg tcgagttctg ccgctcctgg cgtcccgact ggatcgtctg  
2461 ggagccgctg accttcgccg cccgatacg ggcccgggtc accggaacc cgacgcccg  
2521 gatgctgtgg ggtccggacg tcgccaccgg ggcccggcag agcttcctgc gactgctggc  
2581 ccaccaggag gtggagcacc gggaggatcc gctggccgag tggttcgact ggacgctgcg  
2641 gcgcttcggc gacgaccgc acctgagctt cgacgaggaa ctggtgctgg ggcagtggac  
2701 cgtggacccc atccccgagc cgctgcggat cgacaccggc gtccggacgg tgggcatgcg  
2761 gtacgtcccc tacaacggcc cctcgggtgt gcccgctgg ctgttgccgg aaccgaaacg

FIG. 7-25

33/70

2821 tcggcggggtc tgcctgaccc tcggcggttc cagccgggaa cacggcatcg ggcaggctctc  
2881 catcggcgag atgttgacg ccatcgccga catcgacgc gagttcgtgg ccaccttcga  
2941 cgaccagcag ttggtcggcg gtcctgctgc ccacctgctg tccggcaaac gtccgtaccg ccgggttcgt  
3001 gccgatgaac gtcctgctgc ccacctgctg gccacacgtg caccacggcg gccacggcag  
3061 ttggctgacc gccgccatcc aagcagctcc agacctcgg cgcggggctg tcgctcccgg tcgcgggggat  
3121 ggtgcacgcc gacctgctg cacctgctg gggcgatcga gcgggttctc gacgagccgg cgtaccgcct  
3181 gaccgccgag cggatgcggg gacctgctg cggatgcg gaccgacccg tcgcccggccc aggtggtcgg  
3241 cggcgcgag gacctggccg ctgcccgat gacttcacc accacccgga ccgctgatg ccggtcccgg  
3301 catctgtcag gacctggccg cggatgcggg cggatgcg gccacgcggc aggcagccgc gtcgaaccgc  
3361 cgagccgcac ctgcccgat ctgcccgat gacttcacc accacccgga ccgctgatg ccggtcccgg  
3421 aatccacacg ccgactttcc cggatgcggg gacctggcg ggcacgcggc aggcagccgc gtcgaaccgc  
3481 tggacacgac agcacggccg gacctgaact cgtgacccg cgtgacccg gatgcggttc gtcgccgcct  
3541 tcctggtctt cttcacgcac gtcctgtcga ggctcatccc gacggttccc gacgagctac gtgtacgccg  
3601 acggccttga cgccttctg cagaccaccg gacgggtggg gacgggtggg ggtgtcgttc ttctttattc  
3661 tcagcggttt cgtgctgacc tggtcggcg tggtcggcg gggccagcga ctcggtgtgg tcgttctggc  
3721 gcagacgggt ctgcaagctc tcccccaacc acctggtcac cgccttcgcc gccgtggtgt  
3781 tggtcctggt caccgggag gcggtagagc gtgagggcgt gatcccgaac ctctgctga  
3841 tccacgcctg gttcccggcc ctggagatct ccttcggcat caaccgggtg agctgggtcgt  
3901 tggcctgcga ggcgttcttc tacctgtgct tcccgtgtt cctgttctgg atctccggtg  
3961 tccgcccggg gcggctgtgg gcctggggcc cgtggtgtt cgcgcgcatc tgggcgggtac  
4021 cggtggtcgc cgacctcctg ctgcccagtt ccccgccgct gatcccgggg cttgagtact  
4081 ccgccatcca ggactgggtc ctctacacct tccctgcgac gcggagcctg gaggttcatcc  
4141 tcgggatcat cctggcccgc atcctgatca ccggtcgggtg gatcaacgtc gggctgctcc

FIG. 7-26



34/70

4201 ccgagggtgct gttgttcccg gtcttcttcg tcgcctcgct cttcctgccg ggtgtctacg  
4261 ccattctctc gtcgatgatg atccttcccc tggttctgat catcgccagc ggcgcgacgg  
4321 ccgacctcca gcagaagcgc accttcacgc gtaaccgggt gtaggtgtgg ctcggcgacg  
4381 tctccttcgc gctctacatg gtccacttcc tgggtatcgt ctacggggcg gacctgctgg  
4441 gggtcagcca gaccgaggac gccccgctgg gtctcgcact cttcatgac attccgttcc  
4501 tcgagggtctc cctgggtgctg tcgtggctgc tgtacagggt cgtcgagcta cccgtcatgc  
4561 gtaactgggc ccgcccggcc tccgcccggc gcaaaccgcg cacgggaacc gaacagaccc  
4621 cttcccgcg gtaagaagga cggtgcatcg gtgaccacct acgtctggtc ctatctgttg  
4681 gagtacgaga gggaacgagc cgacatcctc gatgcgggtc agaaggctct cgcagtgggc  
4741 agcctgatcc tcggtcagag gcgtgggagc tgtggagaaac ttcgagaccg agtacgcccg ctaccacggg  
4801 atcgcgcaat gcgtgggagc cgacaacggc accaacgctg tgaactcgc gctggagtcg  
4861 gtaggtgtcg gacgcgacga cgaggctcgc acggcttcca acaccggcg cccacagtc  
4921 ctggccatcg acgagatcgg cgcggggcgg cgcccgcccg gtcttcgtgg acgtccgcga cgaggactac  
4981 ctcatggaca ccgacctggt ggaggcgggc gtcaccccg gtaccaaggc catcgtcccg  
5041 gtgcacctgt acgggacagt cgtaggacatg acagccctgc gggaaactggc cgaccggcgg  
5101 ggcctcaagc tcgtggagga ctgcgcccag gccacgggtg cccggcgggg cggtcggcgtg  
5161 gccgggacga tgagcgacgc ggcgcccttc tcgttctacc cgacgaaggc cctcggcgcc  
5221 tacggcgacg gcggcgcggt cgtcaccaac gacgacgaga cagcccgcg cctgcgacgg  
5281 ctgcgggtact acgggatgga ggaggctctac tacgtcacc ggaccccggg tcacaacagc  
5341 cgcctcgacg aggtgcaggc cgagatcctg cggcgcaaac tgaccgggt cgacgcgtac  
5401 gtcgcgggtc ggcgggcggt cgcccagcgg tacgtcgacg ggctcgccga cctccaagac  
5461 tcgcacggcc tcgaactccc agtggtcacc gacggcaacg aacacgtctt ctacgtgtac  
5521 gtcgtccgcc acccgcgccg cgacgagatc atcaagcgtc tccgggacgg gtacgacatc

FIG. 7-27

35/70

5581 tccctgaaca tcagctaccc ctggccggtg cacaccatga ccggcttcgc ccacctcggc  
5641 gtcgcgtcgg ggtcaccgaa cggctggccg gcgagatctt ctcccttccc  
5701 atgtaccctt ccctccctca cgacctgcag gacagggtga tcgaggcggc gcgggaggtc  
5761 atcacccggc tgtgacgagc ccgcgtgtcg tcagcgaaga cccactctgg aagggccggc  
5821 catgccgaac agccactcga ccacgtcgag caccgacgtc gcccgtacg agcgggcgga  
5881 catctaccac gacttctacc acggccgtgg caagggtac cgtgccgaag ccgacgcgct  
5941 cgtggaggtc gcccgcaagc acacccaca ggccggcgacc ctgctggacg tggcctgcgg  
6001 gaccggatcc cacctggtcg agctggcgga cagcttcggg gaggtggtgg gggctcacct  
6061 gtcggccgcc atgctcgcca ccgccgcccg caacgacccc gggcgggaac tgcaccaggg  
6121 cgacatgcgc gacttctccc tcgaccgcag gtctgacgtc gtcacctgca tgttcagctc  
6181 caccggttac ctgctcgacg aggccgaact ggaccgtgcc gtggcgaacc tggccggtca  
6241 cctcgcgcct ggccggcacc tcgtcgtgga gccctggtgg tccccggaga cgttccggcc  
6301 cggctgggtc ggggccgacc tggtcaccag cggtgaccgg aggatctccc gcatgtcgca  
6361 caccgtcccg gcgggtctgc ccgaccgcac cgcctcccgg atgaccatcc actacacggt  
6421 ggggtcaccc gaggccggga tcgagcactt caccgaggtg cactgatga ccctgttcgc  
6481 ccgcgccgcc tacgagcagg ccttccagcg ggccggccctg agctgctcgt acgtcggcca  
6541 cgacctgttc tcgccgggcc ttttcgtcgg ggtcgcgcgg gagccggggc ggtgagggtc  
6601 gaggagctgg gcatcgaggg ggtcttcacc ttcacccccc agacgttcgc cgacgagcgg  
6661 ggggtgttcg gcacggcgta ccaggaggac gtgttcgtgg cggcgtcgg ccgccgctg  
6721 ttcccgggtg cccagggtcag caccacccgg tcccggcggg gtgtggtccg gggggtgcac  
6781 ttcacgacga tgcccggctc catggcggaag tacgtctact gcgccagggg tagggcgatg  
6841 gacttcgccg tcgacatccg gcccggttcc ccgaccttcg gccggggccga gccgggtcgag  
6901 ctctccgccg agtcgatggt cgggctgtac cttcccgtgg gcatggggcca cctgttcgtc

FIG. 7-28

36/70

6961 tccctggagg acgacaccac cctcgtctac ctgatgtccg ccggttacgt cccgacaag  
7021 gaacggggcgg tgcacccctt ggatccggag ctggcgttgc cgatcccggc cgacctcgac  
7081 ctctgtcatgt ccgagcggga ccgggtcgca ccacccctcc gggaggcccc ggaccagggg  
7141 atcctgcccg actacgccg tgcggggcgg tgggtggtgct ctgccgggcc gccgcgcacc gacgtgaccc  
7201 cggccggggc caccacgcc ctggccgacc tcccgggtgcg ggtgcggctc cgtgccccggc ggtcggcggt  
7261 cgtgccctcc agcgtctcgg gaggtggtcg cggacgcccg gacgtcttc ggcggtcctca ccgaacccgg  
7321 agcgtctcgg gaggtggtcg cggacgcccg ggcggtcttc ccgttcgccg ccagatcag  
7381 ggggtacgtca cctgggtccg gacctgacg tccagcgagg tccagcgtggtc gccgaacgga cgaacgtcgg  
7441 cctgggtccg cccggggcagg aacacgcagg tccggcagggt tccggcggc caccgcccg tgggtggtctt  
7501 gcaggaccac cccgagggcg gcggccgggg cgatccggcg ccgacgaccg ggggggtggtc gacgtcgcg  
7561 ggtgcccggc cctgaccggc cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7621 ggaggccact ggaggccggc gggggccggc gggggccggc gggggccggc gggggccggc gggggccggc  
7681 ggtgcccggc cctgaccggc cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7741 ggtgcccggc cctgaccggc cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7801 cctgaccggc cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7861 cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7921 acgccacttc cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac cgtgaccgac  
7981 ctgcgcagc gtcgcccggc atggaccctg cggacctgag cagcgtggag gacccgggtg ccggtggtct  
8041 tccggcgcac cccggcgccc ggggtggcggg ccacgggtcac gatggcgagg gctgaccgcc ggtcacccgg  
8101 ggctgtcacc ggcgttggcc cccggcgccc ggggtggcggg ggtcgtcgac ggtcgtcgac ggtcgtcgac  
8161 ggcgttggcc cccggcgccc ggggtggcggg ggtcgtcgac ggtcgtcgac ggtcgtcgac ggtcgtcgac  
8221 ggtcgtccta cggcaccggc ccgtcgacgg ccgtcgacgg ccgtcgacgg ccgtcgacgg ccgtcgacgg  
8281 ggagttcctc ctcgccagc gtcagctcgg cggcccgtaa cggcgagtcg agctgctcgg

FIG. 7-29

37/70

8341 gtgtgcgggg gccgatgaca gcgcccagga tcccggggcg ggacaggacc cagggcagac  
8401 cgacctcggc cgggtccgcg ccgagggcgtc ggcagtagtc ctctacgcc tcgacgaggg  
8461 ggcgtacggc ggggaggagc acctgggcgc gtccctgcgc cgaactgacg gcggttccgg  
8521 ctgccaaactt ctccagtagc ccgctgagca gccgcctg gcccgcctg caggggggac cagggcgaaca  
8581 cgcccacccc gtacgcctgg gcggcgggca ggacgtccag gcagcagggt tggaggagcc gacgtaccg accttcccac  
8641 ggttgtagcag gactgggtgg gcactgcatgc gagcagggt tggaggagcc gacgtaccg gacgtaccg accttcccac  
8701 gggcgggcggc gatgtgccag cccgccagggt gcctgccaca cctcgtccca cccgaggcg cgtcgttgac ccgttcgctc atctcgtgc  
8761 tgccgaccag atgttcggcg gtagatgtcg atgtggtcga cccgaggcg cgtcgttgac ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
8821 ggtgcgtctg gtagatgtcg atgtggtcga cccgaggcg cgtcgttgac ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
8881 cggcgacgat gtgtcgggcg gagagcccgc gtctcctcgc cccttgtaga gccgccagcc ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
8941 ccaccttggt cggcaggacg gtctcctcgc cccttgtaga gccgccagcc ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
9001 cgacgagttc ctcggtgtgg cccttgtaga gccgccagcc ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
9061 tgcagttgac gcccgctcg agggcgtggt agggcgtggt ccgttcgctc gcccaggcg gtagatgtcg gcggtgtcga  
9121 cccgtccact gaagttcacg gtgcccagcc gtgcccagcc agagtcggct agagtcggct agagtcggct agagtcggct agagtcggct  
9181 cgacgcgtac ccggggcgac ccggggcgac ccggggcgac ccggggcgac ccggggcgac ccggggcgac ccggggcgac ccggggcgac  
9241 gtgctggtgg gcgagcgctt gcgagcgctt gcgagcgctt gcgagcgctt gcgagcgctt gcgagcgctt gcgagcgctt gcgagcgctt  
9301 cgcctcctgc cgcagcttct cgcagcttct cgcagcttct cgcagcttct cgcagcttct cgcagcttct cgcagcttct cgcagcttct  
9361 gagagcctgc cagaggggtgt cagaggggtgt cagaggggtgt cagaggggtgt cagaggggtgt cagaggggtgt cagaggggtgt cagaggggtgt  
9421 cagctcggcg gtgcgctgac gtgcgctgac gtgcgctgac gtgcgctgac gtgcgctgac gtgcgctgac gtgcgctgac gtgcgctgac  
9481 cggtacgccg tggtgcagcg tggtgcagcg tggtgcagcg tggtgcagcg tggtgcagcg tggtgcagcg tggtgcagcg tggtgcagcg  
9541 ggcacagccc ggcagcagga ggcagcagga ggcagcagga ggcagcagga ggcagcagga ggcagcagga ggcagcagga ggcagcagga  
9601 caccgacgcc ggatcgagcc ggatcgagcc ggatcgagcc ggatcgagcc ggatcgagcc ggatcgagcc ggatcgagcc ggatcgagcc  
9661 ggtggccagt gtccggagga actcctgcgg actcctgcgg actcctgcgg actcctgcgg actcctgcgg actcctgcgg actcctgcgg

FIG. 7-30

38/70

9721 cccggtgaag cagaccggc ggactccgtc cgaggtcctg agccactgcg gcacgacgga  
9781 ggaccggttg tagggcaaaag tccgggtgtg caccgactcc agtccggtct ccaggcgga  
9841 gctctcggc agctggtcga cgctccactg tccgacagcg aggtcctcgc ttagtcgag  
9901 gccgaaccg gccgacact ccgtgagcca gccgccgagc ggtccggcc ggtcgtcggc  
9961 gggacgctgc ccgagcagg cctgggagcg gctgcggaag tagccggtga ggtcgtgccc  
10021 ccacagcagc cgggcgtggg cggccccgca ggccttggcc gcgaccgcc ccggaagg  
10081 gaagggtcc cagagcacca ggtcgggacg ccagtccatg gcgaactcga cgagttcgtc  
10141 gacgaaggag tcgttgttga ccaccgggaa gacgaaccgg gaggtggcct cctcgatgcc  
10201 gtgcaggaa tccacgagc gcagtcccg cgcgtcggg gcgaagtcca ggtcgtggt  
10261 gtagcgtgc acctgcgcgg cggcctcagg ggagatgtcg aagagtcggt ggtccgagcc  
10321 gagtggcacc gaggtcagtc ccgcgccgac gacgacgtcg gtgagctcgg gctgactggc  
10381 caccgggacg tcgtggccgg cgtgtgtgag cgtcgtgagg agggggacga ggcctggaa  
10441 gtgggtacgg tgcgcgaacg aggtgagcag gacctgact ggtcactcct tggtcgagat  
10501 gagggcggca acggtccggt cgatgccctc ggccagcggc acccgggggt gccagccggt  
10561 cagcgtccgg aactcgggtg agtcgaagtc gtcgctgagg aagtcgttgg cctcggcgtt  
10621 ctccgggtgga gggacgctga cgacggggcac cgagggttg ccggtctgac gtgccacgct  
10681 ggcggcgacg gtctcgaaga tctcggccgag ggtcggggcc tcgtccgcgc tcggcgtcca  
10741 gacgtcgcgg accagcgcct cgtgggttgtg cagtgcggcg gtgaacgcgg tggccacgtc  
10801 ctcgacgtgc aggaggttgc ggcgcacgct gccctcgtgc cacatcgtga tcggctcacc  
10861 ggcgagggct cgcgggatca tggcgggtgac gacaccccg ccggtctgcc ccgacgggccc  
10921 gctgtggccg tagatcgcgg gcaggcgcag gatcaccccg tcgacgaccc cgtcctcgtt  
10981 ggcctgacgc aggatccgct cggcctcgat ctgtgtgctgg gcgtaccggc tggggggcggc  
11041 ggggttcgcg gcctgggtgg tgctggcgaa caggagcacc ggcgcgggtc cgggtcttgc

FIG. 7-31



39/70

11101 ccgcagcgcg gcgacgaggt cgcgcatgat gccgcggttg acgcgtttcgg cctcggggcac  
11161 cgtggcgggcg ctgcccaggg tcgacccgcc tcgagggcgtag gcgaccagat gcacgacgac  
11221 gtcggtgtcg gcgacgacct gcgacgacct gcggggttcg agcagggtcga ctcgaaaggtg  
11281 ctcgatcccg gcgctgcctg gtggctggtc gcgagacccg gtgcgcgcga cggcccgcag  
11341 tcggagaggg tgtgtggtaa attcgcgaa atcgcatccg aagggcgctt ccgacgaatc cagaaacgcc  
11401 gagaagtgtg acatgtcttg tcatctacta atgcattccg atagccaccg gcgcatggaa  
11461 tccatttgtt cccccaggg tggtgtcggg tgacaaatcc ggcctcaggt cggcctcaag  
11521 cctctttcga gcgggtgctg aggtttcccg cgtaccctcg gtggcctgcyg ttcggggcggg  
11581 tgtcggggaa agggcggatc gaggagtctg gtagggcgctc gcggcgcgta ctccgggact  
11641 gatccgggtc gacgccccga cgctgacag ggctcgatc cgtgccgccc gtaccgccgg  
11701 ttttcggcga tggtcgcaga ttctcccga cgtggtggac tcattggttc tcccgggtgt  
11761 ggccgcaccg tcggtggcct cgtcgggggt gtcggagacc gggtcgatcg ccgtcccccg  
11821 ccgtgccgac cagggtcggc ccgtcgccga ggtgggtcac cgtcgggtgg acccgggtccg  
11881 ccggcgggca ccgcccgatc gtgcccacct tgccctccgc gggtaaatgc ttcgtcgatc  
11941 tgatcgacac ttccggcgac gctatcacg ggcattccc cggcaccacc ggtcgatgcc  
12001 tcgcgcttcc caaacaggga aaacagcagc tcacagcggt tccaggcgcc gggcaatcct  
12061 agcgaagagt ctcgatgggg tcaagggtgaa ttctgtcaca gatgtttttg ttaaatgtac  
12121 tttcttcagc caccctcgac gttcatacaa ttggccggga tctctaccaa gggggagtga  
12181 gtggttgacg tgcccgatct actcggcacc cggactccgc acccagggcc gctcccatc  
12241 ccgtggcccc tgtgcggtca caacgaaccg gagctgcggg ccgcgcccgc tcaattgcac  
12301 gcatactctg aaggcatctc cgaggatgac gtggtggccg tcggcgccgc cctcgcgcgc  
12361 gagacacgcy cgcaggacgg gccgcaccgc gccgtcgctg tggcctcctc ggtcacccgag  
12421 ctgaccgccg cgctcgccgc cctcgcccag ggccgcccac accctcgggt ggtacgcgggt

FIG. 7-32



40/70

12481 gtcgcccagac ccacgggcacc ggtggtgttc gtcctgccc gtcagggcgc ccagtggccc  
12541 ggcatggcga cccgactgct cgccgagtcg cccgtcttcg ccgcggcgat gcgggcctgc  
12601 gagcgggcct tcgacgaggt caccgactgg tcgttgaccg aggtcctgga ctaccccgag  
12661 cacctgcgcc ggcctcgaggt ggtccagccc gcgctcttcg cggtgcagac ctactggcc  
12721 gccctgtggc ggtcgttcgg ggtgcgaccc gacgcccgtac tcggacacag catcggtgag  
12781 ctggccgccg ccgaggtctg cggcgccgtc gacgtcgagg ccgcccgcgc ggccgcgcgc  
12841 ctgtggagcc gcgagatggt cccactggtg ggccggggtg acatggcggc ggtggcgctc  
12901 tccccggccg agctggcagc ccggtcgag cggtgggacg acgacgtcgt gccggccggg  
12961 gtcaacggtc cccggtcggg cgtgctcacc gctgctcacc ggcgctccc agcccatcgc acggcggggtc  
13021 gccgagctgg cggcacaggg cgtacgcgc caggtcgtca acgtgtcgat ggccggcgcac  
13081 tcggcgccagg tcgacgccgt cgccgagggc atgcgctcgg cgctgacctg gttcgccccc  
13141 ggcgactccg acgtgcccta ctacgccggc ctacccggcg gccggctgga caccgggaa  
13201 ctcggcgccg accactggcc gcgcagtttc cggtcccgg tgcgcttcga cgaggcgacc  
13261 cgtgcggtcc tggaaactga gcccggcacg ttcatcgagt cgagcccga cccggtgctg  
13321 gcggcctccc tgcagcagac cctcgacgag gtcgggtccc cgcccgcgat cgtgccgacc  
13381 ctgcaacgcg accagggcgg tctgcgccgg ttcctgctcg ccgtggcgca ggcgtacacc  
13441 ggtggcgtga cagtcgactg gaccgccgc cctcgacgg agttcgactg ggccgcgcgc  
13501 tcggccgctg ccgtcgagac gacgagggg taacccgggg tgaccccccg ccacctgccc  
13561 gaccacgtac tgcgcgcgcg gctgctggag atcgtcggcg ccgagacggc cgcgctcgcc  
13621 gggcgggagg tcgacgcccg ggccacctc cgggaaactgg gcctcgactc ggtcctcgcg  
13681 gtgcagctgc ggacccgcct cgccacggcg accgggccc atctgcacat cgccatgctc  
13741 tacgaccacc cgaccccgca cgccctcacc gaggcgctgc tgcgcggccc gcaggaggag  
13801 ccggggcggg gtgaggagac ggcacacccg acggaggccg aaccgacga acccgtcgcc

FIG. 7-33

41/70

13861 gtggtcgcca tggcgtgccg gctgcccggc ggcgtcacct caccggagga gttctgggag  
13921 ctgctggccg aggggaggga cgccgtcggc gggctgccc cgcaccgggg atgggacctg  
13981 gactcgctgt tccaccggga cccgaccgg tccggcacgg cgcaccagcg cgctggtggc  
14041 ttcctcacgg gcgccacctc cttcgacgct ccttctctcg ggctgtcgcc acgggaggca  
14101 ctggccgctg agccgcagca gcggatcacg ttggagctgt cgtgggagggt gctggaacgc  
14161 gccgggatcc ccccgacgtc gttgcggacc gttgcggacc tcccggaccg ggggtgttcgt cggctctgatc  
14221 cccaggagt acggcccccg gctggccgag gggggtgagg gcgtcgaggg ctacctgatg  
14281 accgggacca ccaccagcgt cgcctccggt cgggtcgcct acacctcgg cctggagggg  
14341 ccggcgatca gcgtcgacac cgcctgctcg tcgtcgtcgtcgcgtgcgtgca cctggcggtgc  
14401 cagtcgctgc ggcgcgggca gtcgacgatg gcgctcgccg gtggcggtgac ggtgatgccg  
14461 acaccgggca tgctcgtgga cttcagtcgg atgaactccc tcgcccccg tgcacgggtcc  
14521 aaggcgttct cggccgccgc cgacgggttc ggcattggccg aaggcgacag gatgctcctg  
14581 ctggaaacggc tctcggacgc ccgccgccac ggccacccgg tgctcgccgt gatcaggggg  
14641 accgctgtca actccgacgg cgcgagcaac ggactctccg cccgaaacgg ccgggccccag  
14701 gtccgggtga tccgacaggc cctcgccgag cctcgccgctga cgccccacac cgtcgacgtc  
14761 gtggagaccc acggcacccg accccgcctc ggtgatccga tcgaggcacg ggcgctctcc  
14821 gacgcgtacg gcggtgaccg tgagcacccg ctgcggatcg gctcggtcaa gtccaacatc  
14881 gggcacaccc aggccgccgc cggtgtcgcc ggtctgatca aactggtgtt ggcgatgcag  
14941 gccggtgtcc tgccccgcac cctgcacgcc gacgagccgt caccggagat cgactggtcc  
15001 tcggggcgcg ttagcctgct ccaggagccc gctgcctggc ccgccggcga gcggccccgc  
15061 cgggccgggg tgctcctcgt cggcatcagc ggcaccaacg cacacgcgat catcgaggag  
15121 gcgccgccga ccggtgacga caccgaccc gaccggatgg gcccggtggt gccctgggtg  
15181 ctctcggcga gcaccggcg ggcgttgccg gcccgggcgg cgcggtggtgc cgggcaccta

FIG. 7-34

42/70

15241 cgcgagcacc ccgaccagga cctggacgac gtcgcctact cgctggccac cggtcggggc  
15301 gcgctggcgt accgtagtgg gttcgtgccc gccgacgcgt ccacggcgct gcggatcctc  
15361 gacgaactcg ccgccggtgg atccggggac gcggtgaccg gcaccgcccg cccccgcag  
15421 cgcgtcgtct tcgtcttccc acccgggtctt cgcctcgggtg tggcagtggg cgggatggc agtcgacctg  
15481 ctcgacggcg tacctggact tcgagatcgt cccgttcctg cgggccgagg gcagcgccg gacccccgac  
15541 cacacgctct ccacgaccg ccggtggcg gccgtacggg gtggaaccgg cgccgctcat cggacactcc  
15601 tccctggcgg cagggggaga ttgccgcggc ggtcgtgccc accatgcccg gcaacggcgc gatggcctcg  
15661 cagggggaga gcggtggccc atcgccgcct gtcaacggtc cgcctcctgca cgcgtcgcac  
15721 gcgggtggccc gcggtggcgc cgcgtcgacc ggtggcggcc ggtggtctcc aagcggctgc cggcgggac  
15781 gcggtggccc gcggtggcgc cgcgtcgacc ggtggcggcc ggtggtctcc aagcggctgc cggcgggac  
15841 atcgccgcct gcaacggtc cgcgtcgacc ggtggcggcc ggtggtctcc aagcggctgc cggcgggac  
15901 gtcaacggtc cgcgtcgacc ggtggcggcc ggtggtctcc aagcggctgc cggcgggac  
15961 gcctcctgca tcctcgacg ctgcccgggct ctgcccgggct ctactcgaca gtcaccggcc ctcacgccc  
16021 tcctcgacg ctgcccgggct ctgcccgggct ctactcgaca gtcaccggcc ctcacgccc  
16081 ctgcccgggct ctgcccgggct ctactcgaca gtcaccggcc ctcacgccc  
16141 ctgacgccc ggtactgggtt tcgcaacctg gtacacgacg ttcctggagg cgtggcgggtg acctcgtcgc tgtccactcg  
16201 cgctccctcg ccgaccaggg tcgagggagt gggccggcgg tcccgtcgac ttccggctccg cgctggcccg cgccttcgtg  
16261 accacggcga ctgcgacgtg gccggcgtcg cagtggaactg acccgttcca gcgtagagcg ttctgggttg aaccgaatcc ggcccgcagg  
16321 ctgcgacgtg gccggcgtcg cagtggaactg acccgttcca gcgtagagcg ttctgggttg aaccgaatcc ggcccgcagg  
16381 gccggcgtcg cagtggaactg acccgttcca gcgtagagcg ttctgggttg aaccgaatcc ggcccgcagg  
16441 ctgcccacgt gtcgccgact ccgacgacgt ctgctccctg cggtagcgca cggtagcgca cccgaccgat  
16501 gtcgccgact ccgacgacgt ctgctccctg cggtagcgca cggtagcgca cccgaccgat  
16561 ccgggtgagc cgggacgggt cgacggcacc tggctgctgg cgacgtaccc cggtcggggc

FIG. 7-35

43/70

16621 gacgaccggg tcgaggcggc gcggcaggcg ctggagtccg ccggggcgcg ggtcgaggac  
16681 ctggtggtgg agcccggac gggccgggtc gacctggtgc ggccggctcga cgcgtgggt  
16741 ccggtggcgg gcgtgctctg cctgttcgct ggacacgctc gacctgacc ggcgggccga aactccccg  
16801 ctggcggtga cgtcgttgct cgatctgggt ggtcacccgag aacgccgtcg ccgtcgggcc cttcgaacgg  
16861 cgggagtgtc ctcgcgacc cggcccacgg cgcgctctgg gccctcggtc gggtcgtcgc cctggagaac  
16921 ctcgcgacc ccgcccgtct ggggcggcct ggtcgacgtg ccgtcgggtt ccgtcggcca gctgtcgcgt  
16981 ccgcccgtct cacctcggga cacctcgtc cggcgccggc gaggaccagg gcgggcggca cgggccctcg accgacggg  
17041 acgtacgccc gcggtggtg gccggtggtg cagggcgggc cagggcggca ggggcggtc ggcacgtcgc ccgtggctg  
17101 ggcaagggtc gcccggcagg gaccccgtg cctggtgctg actcggccgac ctccgtcga ccgtcgacga gacctcacc  
17161 ggcacgggtc ggcacgggtc tactcacga accgggagca gctccgtgcc cgcggcgacg ctcgacgacg aaggctctc  
17221 gcccggcagg gtcgaggagc gacgtcaccg tggtccacgt tcgaacgggc caaccgggcg aaggtgctcg gtgcccgcga  
17281 gtcgaggagc gacgtcaccg accgggagca gctccgtgcc cgcggcgacg ctccgtcga ccgtcgacga gacctcacc  
17341 gacgtcaccg accgggagca gctccgtgcc cgcggcgacg ctccgtcga ccgtcgacga gacctcacc  
17401 ctgtcggcgg ggtgaccgca ctgacccggg agcgggagca ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17461 ggtgaccgca ctgacccggg agcgggagca ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17521 ctgacccggg agcgggagca ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17581 ggcgcggcgg ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17641 cagcgacgca ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17701 gggatggccg ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17761 ccgaccagg ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17821 gtcgtcgaca ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17881 ctcttcgaca ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc  
17941 gtggcgggcg ggcgcggcgg ggcgtcggcg gtgcccgcga cctcctccac cgcgcgctc

FIG. 7-36

44/70

18001 cggacgcacg cggctgccgt cctcggccac gcctcggccg agcaggtgcc cgtcgacagg  
18061 gccttcgccg aactcggcgt cgactcgctg tcggccctgg aactgcgcaa ccggtgacc  
18121 actgcgacccg gggctccggct ggcacgacg acggtcttcg accaccgga cgtacggacc  
18181 ctggcccgac acctggccgc cgaactgggc ggcggatcgg ggcgggagcg gcccgggggc  
18241 gagggcccca cggtgggccc gaccgacgag ccgatacgcca tcgtcgggat ggcctgccgg  
18301 ctgccggggg gactggactc accggagcag ctgtgggagt tgatcgtctc cgggcgggac  
18361 accgcctcgg cggcacccgg ggaccggagc tgggatccgg cggagttgat ggtctccgac  
18421 acgacgggca ccgtaaccgc cttcgggcaac ttcataggccg gggcgggca gttcgacgcg  
18481 gcgttcttcg ggatctcggc gcgtgaggcg ttggcgatgg atccgcagca gcggcacgcc  
18541 ctggagacca cctgggaggg cctggagaac gccggtatcc gcccagatc gttgcgcggt  
18601 acggacaccg gtgtcttcgt gggcatgtcc catcaggggt acgccaccgg ccgcccgaag  
18661 ccgaggacg aggtcgacgg ctacctgtg acaggcaaca ccgcgagcgt cgcctccggt  
18721 cggatcgcgt acgtgttggg gttggagggg ccggcgatca ctgtggacac ggcgtgttcg  
18781 tcgtcgcttg tggcgttgca cgtggcggcg ggttcgttgc gttctgggga ctgtggtctg  
18841 gcggtggcgg gtggggtgtc tggctccgga cggcaggtgc aagccctct tgacagcgg  
18901 cagggcgcgt aggggtcggc cttcgtcgtg ttgcagcggg cggacgagc cgcggcttc  
18961 ggtctggggg ggtcgtcggg tgttgggtgt cggcggtga atcaggatgg ggcgagtaat  
19021 gggcgtcggg cgcctcggg ggtggcgag cagcgggtga ttcggcgggc gtggggtcgt  
19081 ggggtggcgg ggggtgggga tgtgggtgtg gtggaggcgc atgggacggg gacgcggtg  
19141 gcgggtgtgt cgggtggggg tggagtggg ggcgttgtg gggacgtatg ggtggggtg  
19201 ggggatcccg tggagtggg tggagggtg ggtgaaggcg aatgtgggtc atgtgcaggc  
19261 ggtccgggtg tggagggtc ggtgggtcggg ttgggtcggg ggttgggtgg tccgatgggtg  
19321 gtggtgggtg

FIG. 7-37



45/70

19381 tgtcggggtg ggttgtcggg gttggtggat tggtcgtcgg gtgggttgggt ggtggcggat  
19441 ggggtgcggg ggtggccgggt ggggtgtggat ggggtgcgtc ggggtgggggt gtcggcgtttt  
19501 ggggtgtcgg ggacgaatgc tcatgtggtg gtggcggagg cggcgggggtc ggtggcgggg  
19561 gcggaacggc cggaggagg gtcgtcgcgg ggggtgggtg ggggtgggtg tgggtgtggtg  
19621 ccggtggtgc tgtcggcaaa gaccgaaacc gacctgcacg cccaggcacg tgcactcgcc  
19681 gaccacctgg agacgcacc cgacgtccc cgacccgacg atgaccgacg tgggtgtggac gctgacgcag  
19741 gcccgccaac gcttcgacag gcgcgcggtc ctctcgccg ccgaccggac ccaggccgtg  
19801 gaacggctgc gcggccctcg gcggggcgaa ccggggaccg gtgtggtgtc ggggtggcg  
19861 tcgggtggtg gtgtggtgtt ggtttttcct tggttttctt gtcagggtg gtcatgtgggt ggggatggcg  
19921 cgggggttgt tgtcggttcc ggtgttgtg ggtgttgtg ggtcgggtg tggagtgtga tgcgggtggtg  
19981 tcgtcgggtg tgggggtttc ggtgttgtg ggtgttgtg gtgttgtggg gtgcgtcggg tgcgccgtcg  
20041 ttggatcggg tggatgtggt gcagccgggt gcagccgtg ttgttcgtg tgatgggtgtc gttggcgcg  
20101 ttgtggcgggt ggtgtgggggt tgtgcctgcg gcggtgggtg gtcattcgca gggggagatc  
20161 gcggcgggcg tggcggggc ggtgttgtc ggtgttgtc gtgggtgatg gtgcgcgggt ggtggcgttg  
20221 cgggcgcggg cgttgccggc gttggccggc ctcgacagc ggcccctgga cggggaagct ggagatcgcc  
20281 cgcgacgacg tacagaagct gcccgcgcg ggtggtggtc cgggttccgg gaccgagctg  
20341 gcgggtcaac gtcgagcact gtgacgggat cggggtccgg gcccggacga tcccgtcga ctacgcctcc  
20401 cactccgcac aggtcgagtc gctccgggag gagctgctct cgtcctggc cgggatcgag  
20461 ggccgccccg gacgggtgcc gttctactcc accctcacc gtgggttcgt cgacggcacc  
20521 gaactggacg ccgactactg gtaccgcaac ctgcgccacc cgtgcgggtt ccacgccgcc  
20581 gtcgagggcg tggcagcgcg tgacctcacc acgttcgtcg aggtcagccc gcaccccggtg  
20641 ctgtcgatgg cggtcgggga gacgcttgcc gacgtggagt ccgccgtcac tgtgggcacc  
20701

FIG. 7-38



46/70

20761 ctggaacgcg acaccgacga cgtcgagcgc ttctcacct ccctcgccga ggcgcacgtc  
20821 cacggcggtac ccgtggactg ggcggcggtc ctcggtctccg gaaccctggt cgacctgccc  
20881 acctatccct tccagggacg gcggttcttg ctgcacccc accgtggtcc gcgtgacgat  
20941 gtcgccgact ggttccaccg ggtcgactgg acggcgacgg ccaccgacgg gtcggcccga  
21001 ctcgacggtc gctggctggt ggtcgtaacc gaggggtaca cggacgacgg ctgggtcgtg  
21061 gaggtgcggg ccgccctcgc cgcgggtggt gccgagccgg tggtagacac ggtcgaggag  
21121 gtcaccgacc ggttcggtga cagcgacgcg gtggtgtcga tgctcgggct ggcgacgac  
21181 ggtgcggccg agaccctggc gctgctgcga cgaactcgac cacaggcgtc caccaccca  
21241 ctgtgggtgg tcaccgtggg ggccgtcgc ccgcccgtc cggtagcagc cccgaacag  
21301 gcgacggtgt gggggttggc ccttgtcgc ccttgtgaac gcggacacgg gtggaccggc  
21361 ctgctggatc tgccgcagac accggaccgg cagctacgac ccgggctggt cgaggcgctc  
21421 gccggtgccg aggaccagg agcggtccgc gccgacgcc gccgacgat cctcgtcac  
21481 cccaccccg gggtcagccg tgccgtcacc gcccgatggc tcgccgagcg cgtgcccga  
21541 gggggcaccc cgggtctggg ccggtctggg gcgcgggccc ggcaccgccc ggtggtccc  
21601 cacctcgccc tggtcagccg gcgcgggccc ggcaccgccc ggcaccgccc ggtggtccc  
21661 gacctgaccg ggctcggcgt acgggtgtcg gtgcaactct gcgacgtcgg cgaaccgag  
21721 tcggtcggcg ccctgggtga cctgggtgca ggagttaga gcagccgggt acgtggtccg gggggtggtc  
21781 cagctgccc cgtgcccc gtctgcccc gaggtgcca ctgaccgaca tggacccggc cgacctcgc  
21841 gacgtggtgg ccgtgaaggc cgacggcgcc gtcacactgg cgcacctgtg cccggaggcc  
21901 gaactgttcc tgctgttctc tgctgttctc ctccggggcc ggggtgtggg gcagtgcccg tcaggggtcg  
21961 tacgccgccg gaaacgcctt cctggacgcc cctggacgcc ttcgcccgac accggcggga cccgggtctg  
22021 cccgccacct cgggtggcgtg ggggctcttg gcggccgggg ggatgacagg ggaccaggag  
22081 gcggtgtcgt tcctgcgtga gcggggcgta cggccgatgt cggtagccgag ggcactggaa

FIG. 7-39

47/70

22141 gcgctggaac gggtcctcac cgccgggggag accgcggtgg tcgtgccga cgtcgactgg  
22201 gcggccttcg ccgagtcgta cacctccgcc cgccccggc cgctgctcca ccggctcgtc  
22261 acacctgcgg cgccggtcgg cgagcgcgac gagccgcgtg agcagaccct ccgggaccgg  
22321 ctggcgggccc tgccccggc cgagcgtcg gcgagactgg tacgcctggt ccggcgggac  
22381 gccgcagccg tgctcggcag cgacgcgaag gccgtaccg ccaccacgcc gttcaaggac  
22441 ctcggttcg actcgctggc cgcggtccgg ttccgtaacc ggctggccgc ccacaccggt  
22501 ctgcgtctgc cggccaccct ggtcttcgag caccgaacg ccgcagccgt cgccgacctc  
22561 ctccacgacc gactcggcga ggccggcgag ccgacccccg tcgggtcggg gggcgccgga  
22621 ctggccgcgc tggagcaggc cctgcccgc cctccgaca cgagcggggt cgagctggtc  
22681 gagcgcctgg aacggatgct cgccgggctc cgcccgagg ccggagccgg ggccgacgcc  
22741 ccgaccgccg gtgacgacct gggggaggcc ggcgtcgac aactcctcga cgcgctcgaa  
22801 cgggaactcg accccaagtg aacccgaact gaccgcagcc gcagccgaag cagagaccga  
22861 ggacctgtga ctgacaacga caagtgggc gagtacctcc gtcgtgcgac gctcgacctg  
22921 cgggccgccc gcaagcgctg gcgcgagctg caatccgacc cgatcgcggt cgtcggcatg  
22981 gcctgccgcc taccgggcgg ggtgcacctc ccgcagcacc tgtgggacct cctgcgccag  
23041 gggcacgaga cggtgtccac cttccccacc gggcgcggtt gggacctggc ccggctcttc  
23101 cacccgacc ccgaccacc ccgaccacc gttcttcggg tacgtcgacc ggtcgacgac  
23161 gtggcggggt tcgacgccga gttcttcggg atctccccg gcgaggccac ggccatggac  
23221 ccgcaacagc ggctgctgtt ggagaccagt tgggagctgg tggagagcg cgcatcgat  
23281 ccgcaactccc tgctgggcac ccgaccggc gttctcctcg gctgggctac gctcggctac  
23341 ggcgagaacg gcaccgaagc cggtgacgcc gagggctatt cggtgaccgg ggtggcacc  
23401 gctgtcgcc cggggcggat ctctacgc cggtggctgg aggggtccgt gatcagcgtg  
23461 gacaccgct gctcgtcgtc gttggtggcg ctgcacctgg cgtcgagtc gctgcggctg

FIG. 7-40

48/70

23521 ggcgagtcga gtctcgctgt cgtcggcggg gcggcgggtca tggcgacacc aggggtgttc  
23581 gtcgacttca gccgccagcg ggcgttgccc gctgacggca ggtcgaaggc cttcggggcc  
23641 gccgccgacg ggttcggctt ctccgagggg gtctccctcg gtctgctcga acggctctcc  
23701 gaggccgaaa gcaacgggcca cgaagtgtg gctgtcatcc gtggctccgc cctcaaccag  
23761 gacggggcca gcaacgggtt cgcgcgccg aacgggaccg ccagcgcaa ggtgatccgg  
23821 caggcgctac gaaactgcgg cctgacccc gccgacgtgg acgccgtgga ggcgcacggc  
23881 accggcacca cgctcggcga ccgatacgag gccaacgccc tgctggacac ctacggccgt  
23941 gaccgggata cggaccaccc gctgtggctg ggtcgggtga agtcgaacat cggccacacg  
24001 caggcggcgg cgggcgtcac cgggctgctc aagatgggtg tggcactgcg ccacgaggaa  
24061 ctgcccgcga ccctgcacgt cgacgagccc accccgcacg tggactggtc ctcgggagcg  
24121 gtacgcctgg cgaccggggg ccggccgtgg cggcgggggtg accggccgag gcggggccggg  
24181 gtgtcggcgt tcggcatcag cgggaccaac gcccacgtga tcgtcgagga ggcacccgag  
24241 cggaccaccg agcgcaccgt cgccggcgac gtcggcccgg gtccgctcgt ggtgtccgcc  
24301 cggtcggcgg cggcgctacg ggcccaggcg gcccaggtcg ccgagctggg gaggggctcc  
24361 gacgtcgggc tggcggaggt cgggcggagc ctggccgtga ccggggcgcg acacgagcac  
24421 cgggcggcgg tggcggcgtc gaccggggcc gaggcgggtg gggggctgcg cgaggctcgcg  
24481 gcggtcgaac cgcgcggcga ggacaccgtc accgggggtcg ccgagacgtc cgggcgcacc  
24541 gtcgtcttcc tcttcccggg acaggggtcc cagtgggtcg ggtggggcg gatgctgctg  
24601 gactcggcac cggcgttcgc cgacacgata cgcgcctgcg acgaggcgat ggcaccgtg  
24661 caggactggt cggtctccga cgtgctccgg caggagccgg gggcacccgg actggaccgg  
24721 gtcgacgtgg tgcagccggt gctgttcgcg gtgatgggtg cgttggcgcg gttgtggcag  
24781 tcgtacgggg tcacccccgc tgcgggtggtg gggcactcgc agggggagat cgccgccgcc  
24841 cacgtggcgg gtgcgctctc cctcgccgac gcggcgaggc tgggtggtgg ccgcagccgg

FIG. 7-41

49/70

24901 ttgctgcggt cgctgtccgg gggcggcggc atgagcgccg tcgcgctcgg tgaggccgag  
24961 gtacgccgcc gactgcggtc gtggaggac cggatctccg tggccgccgt caacggaccc  
25021 cggtcggtgg tggtggccgg ggaaccggag gcgctgcggg agtggggacg ggagcgggag  
25081 gccgaggggc tacgggtccg cgagatcgac gtcgactacg cctcgcactc gccgcagatc  
25141 gacaggggtcc gtgacgaact cctgacggtc cctgacggtc acgggggaga tcgagccccg gtcggcggag  
25201 atcaccttct actcgacggt cgacgtccgt gctgtcgacg gacccgacct ggacgcgggg  
25261 tactgggtacc gcaacctgcg ggagacggtc cggttcgccg acgcgatgac ccggttggcc  
25321 gactcgggat acgacgcgtt cgtcgaggtc agcccgcac cggtggtggt gtcggcggtc  
25381 gccgaggcgg tcgaggaggc aggtgtcgag gacgccgtcg tcgtcggcac cctgtcccgg  
25441 ggcgacggcg gaccgggggc gtccctgcgg tcggcggcca cggccactg cgcgggtgtg  
25501 gacgtcgact ggacgcccg cctccggga gctgcgacga tccgttgcc gacgtaccg  
25561 ttccaacgga agccgtactg gctgcggtcg tctgctccc ccccgcctc ccacgatctc  
25621 gcctaccggg tgctcctggac gccgatcac ccgcccggg acggcgctact cgacggcgac  
25681 tggctggtgg tgaccccccg tgaccccccg ggaagggtcg acgggttggc ggcggcgatc  
25741 accgccggcg gtggccgggt cgtcgcccac cgggtggact ccgtgacctc ccggaccggc  
25801 ctggccgagg cgctcgcccg gcgggacggc acgttccggg gggtgctgtc gtgggtggcg  
25861 accgacgaac ggacagtcga ggccggtgcg gtcgccctgc tgacctggc gcaggcgttg  
25921 ggtgacgccg gaatcgacgc accactgtgg accctgacct tgcctgacct ccgtaccccc  
25981 gtcgacgggtg acctggcccc accggcgcag gccgccctgc acggtttcgc ccaggtcggc  
26041 cggctggagc tggcccccg cggtcggtgg gtgctcgacc tgccccccac cgtcgacgcc  
26101 gccgggacgc gtctggtcgc ggcggtcctc gccggcggcg gcgaggacgt cgtcgccgtc  
26161 cgtggcgacc gtctctacgg ccgtcgccctg gtcagggcgga ccctgccgcc gcccgcgggg  
26221 gggttcaccc cgcacggcac cgtcctggtc accggcgcgg ccggtccggg gggcgggtcgg

FIG. 7-42

50/70

26281 ctggcccggt ggctgccga acggggtgcc accgactcg tcctgccgg cgcacccg  
26341 ggcgaggagt tgctgaccg gatccgggc gccggtgga ccgctggt gtgcgaaccg  
26401 gaggcggagg cactgcgtac ggcgatcggc ggggagttgc cgaccgcgt cgtacacgcc  
26461 gagacgttga cgaacttcgc cggcgtcggc gacgccgacc ccgaggactt cgcgccacc  
26521 gtcgcggcga agaccgcgt tctactgctc gccgacggtc ctggcggagg tgctcggcga ccaccgcctc  
26581 gaacgggagg tctactgctc gtcggtggcc ggggtctggg gtggggtcgg catggcccg  
26641 tacgccgccg gcagcgccta cctcgacgcc ctggtcgagc accgtcgcgc ccgggggcac  
26701 gccagccct cggtggcctg gaccccggtg gacctgccg gcgcggtcga cgacggtcgg  
26761 ctgcgcgagc gcggcctgcg cagcctcgac gtggccgacg ccctcgggac gtgggaacgt  
26821 ctgctccgcg ccggtgcggt gtcggtggcc gtcgccgacg tcgactggtc ggtcttcaca  
26881 gaggtttcg cggccatccg gccgaccccg ctcttcgacg aactcctga ccggcgccgg  
26941 gaccccgacg gcgcgccct gcacggcccg ggggagccgg cgggcgagt gggtcgacga  
27001 atcgcggcg tgctcccga ggaacagcg ggaacgttg gagacctg tgacctcgt cggcgagacg  
27061 gtcgcggagg tgctgggaca cgagaccggc accgagatca acaccgtcg ggccttcagc  
27121 gaactcggcc tcgactcgct gggctcgatg gccctgcgtc agcgcctggc ggcccgtaac  
27181 ggcctgcgga tgccggcctc gctggtctta gaccaccga cggtcaccgc gctcgcgcgg  
27241 tacctgcgtc gactggtcgt cggggactcc gaccgaccc cggtaagggt gttcggcccc  
27301 accgacgagg ccgaaccct cgcctggtc ggcacgggt gccggttccc cggcggcatc  
27361 gccaccccg aggacctctg gcgggtggtg tccgagggca cctccatcac caccggattc  
27421 ccacccgacc ggggctggga cctccggcgg ctctaccacc ccgacccgga ccaccccgcc  
27481 accagctacg tcgacagggg gggattcctc gacggggccc cggacttcga ccccggttc  
27541 ttcgggatca cccccgca ggcgctggcg atggaccgc agcagcggct caccctggag  
27601 atcgcggtgg aggcggtgga acgggggggc atcgaccgg agaccctct ccggcagcgac

FIG. 7-43



51/70

27661 accggcggtct tcgtcggcat gaacggccag tcctacctgc aactgctgac cgggaggggt  
27721 gaccgggtca acggctacca ggggttgggc aactcggcga gcgtgctctc cggccgtgtc  
27781 gcctaacct tcgggtggga ggggcccggc ctgacggtgg acaccgcctg ctgctcctcg  
27841 ctggtcgcca tccacctgc ccgacggcg gtgcaaggcg ttctccgcgc aggcgacgg gttcgccctc  
27901 gccggcgggg tgacggtcat ggccgacccg gtgcaaggcg ttctccgcgc aggcgacgg gttcgccctc  
27961 gggctcgccg ccgacggcg tcgcccgcgt cgtcctcgaa ccgttgtcca aggcgcggcg aaacggccac  
28021 gccgagggcg cggcggcgt cggcagcgcc cgtcaaccagg acggggccag caacggcctc  
28081 cagggtgctgg acgggcccgc cgggaaagg gtgatacagg aggcctgac cgcctccggg  
28141 gccgccccga acgggcccgc ccgacgtcga catggtggag gcgcacggga actcggcgac  
28201 ctgctgcccg ccgacgtcga catggtggag gcgcacggga actcggcgac  
28261 ccgacgtcga catggtggag gcgcacggga actcggcgac  
28321 ctgggctcgg tgaagacgaa catcggccac catcggcccg acccaggcgtg cgcgggggtg  
28381 atcaaggcgg tcctggcgat gcggcacggc gttactcccga ggtcgtgca cgcgcgacgag  
28441 ttgtccccgc acatcgactg ggcggacggg aaggtcgagg tgctccgcga ggcacgacag  
28501 tggccccccg gtgagcgccc ccgcccgcgc ccgcccggc ggggtgtcct cctcggcgt cagcgggacc  
28561 aacgcccacg tcacgtcga ggaggaccc ctcgtcctg caccgacgca acccgaacc ggttcccgc  
28621 gccccggggc ggcccctgcc ccctcgccga acacctgcgc acccggacct cgcggcacc  
28681 caggcgcgga cctcgccga tggccaccgg tcgcccctg cgccctcgac aggatcgccc ctgcccgcac  
28741 gcccgtaacc tggccaccgg tcgcccctg cgccctcgac aggatcgccc ctgcccgcac  
28801 gaccgggagg gtgtctgcgc cgccctcgac aggatcgccc ctgcccgcac  
28861 gtcgtcgccc cggcggtctt cggcgcccgt acccccgctc tggctctccc cgggcagggg  
28921 tcgcagtggg tcggcatggc ccgtgacctg ctgacctc cagaggtgtt cgcgaggtcg  
28981 atgggcccgg gcgcccagg gctgtcgccg tacaccgact gggacctgct cgacgtgggt

FIG. 7-44



52/70

29041 cgtggggtcg gcgaccccgga ccgtacgac cgggtggacg tgctccagcc ggtgctgttc  
29101 gcggtgatgg tgctgctggc gcggttgtgg cagtcgtacg ggtgactcc ggtgcggtg  
29161 gtgggtcact cgagggggga gatcgccgcc ggcacgtgg ctggtgcgtt gtcgttggcc  
29221 gacgccgcca ggtggtggc gttgcgcagc cgggtgctgc gggagctcga cgaccagggc  
29281 ggcatggtgt cggtcggcac ctcccgcgcc gagtggact cgttggtggc cggaccagac  
29341 gggcgggtcg cggtgggcgc ggtgaacgga ccgggcacgc cgttggtggc cggaccagac  
29401 gccgaactgg acgagttcct cgcggtggcc gagggcccgcg agatgaggcc gcgtcggatc  
29461 gcggtgcgct acgcgtcgca ctcccggag gtggcccggg tcgaacagcg gctcgcgcc  
29521 gaactcggca ccgtcacccg cgtcggcgcc acggtcccgc tctactccac cgccaccggg  
29581 gacctcctcg acaccacagc catggacgcc ggtactggt accgcaacct gcgccaaccg  
29641 gtgctgttcg agcacgccgt ccgcagcctc ctggagcggg gattcgagac gttcatcgag  
29701 gtcagcccgcc accctgtgct gctgatggcg gtcgaggaga ccgccgagga cgcgagcgc  
29761 ccggtcaccc gcgtgccgac gctgcgccgc gaccacgacg gcccgctcga gttcctccgc  
29821 aacctcctgg gggcgcacgt gcacggggtc gacgtcgacc tgcgctccggc ggtcgcgcc  
29881 ggccgcctgg tcgacctgcc cacctacccc ttcgacaggc agcggctctg gcccaagccg  
29941 caccgcaggg ccgacacctc gtcgctgggg gtcctgact cgacccaccc gctgctgcac  
30001 gccgcagtcg acgtacccgg tcacggcgga gcggtgttca ccgggcggct ctccccgac  
30061 gagcagcagt ggctgaccca gcacgtgggtg ggtgggcgga acctggtgcc cggcagtgtc  
30121 ctggtcgacc tcgcgctcac cgccggggcc gacgtcggcg tgccggtgct ggaggaaactc  
30181 gtcctgcagc agccgctggt gttgaccgcc gccggtgcgt tgctgcgcct gtcggtcggc  
30241 gccgccgacg aggacggggc gcggccgggtc gagatccacg ccgccgagga cgtctccgac  
30301 ccggccgagg ccggtggtc ggcgtacgcg accgggaccc tcgccgtcgg cgtggccggc  
30361 ggccggccgg acggcacaca gtggcccccg ccggcgcca ccgccctgac gttgaccgac

FIG. 7-45

53/70

30421 cactacgaca ccctcgccga actgggctac gaggtagggc gagtagggc cggcgttcca ggcgtgcgc  
30481 gccgcgtggc agcacggcga cgtggtctac gcggaggtgt ccctcgacgc cgtcagaggag  
30541 gggtagcgt tcgacccggt gctgctcgac gctgctgccc gccgtgccc agaccttcgg cctgaccagt  
30601 cgcgccccc cgcgagcctcc ctctgcctgg cttgcctgg cggggcgctca ccctgcacgc caccggggcc  
30661 actgcggtac gggtaggtggc gacccccgc gccacggacg ggaccggacg cggtagccct gcgggtcacc  
30721 gaccgaccg gtcagctcgt gccacgggtg cgcacgggtg gacgacctgg gacctgcacc tcgtcagggg cgcggggggc  
30781 gatcgggacc agccgcgcgg ccgcgacggc ccgcgacggc gacctgcacc gcctggagtg ggtacgggctg  
30841 gccaccccg gccaccccg acccgacccc ggcggcggtg gtgcacgtgg cggccgacgg ggtcgcacgac  
30901 ctgctgcgcg ccggtggtcc ggcaccacag gccgtcgtcg tccgctaccg tcccgacggc  
30961 gacgaccga cggccgaggg ccgtcacggg gtgctctggg cggccacgct cgtgcggcgt  
31021 tggctcgacg acgaccggtg gcccgccacc accctggtgg accctacgct cgcagggggtc  
31081 gaggtctccc cgggggacga cgtgccgcgc cgtgctctgg cccggggcgg cggcgtgtg  
31141 cgctgcgccc aggcggagtc cccggaccgc ttcgtgctcg tcgacggcga cccggagacg  
31201 cccccggcgg tgcgggacaa tccgcagctc gcggtccgtg acggtgcggg gttcgtgcc  
31261 cggctgacgc cgctcgccgg tcccgtagcc cccgtcgccc gccgtcgccc accggggtg  
31321 cccggcaaca gcggctccat cgaggcagtg gccttcgccc cgttcccga cgcgacccg  
31381 cccctggcgc cggaggaggt acgcgtcgcc gtccgcgcca cggcgtgaa cttccgtgac  
31441 gtcctgctcg cgctcgccat gtaaccggaa ccggccgaga tgggcaccga ggcgtccggt  
31501 gtggtcaccc aggtcgggtc gggtagtcgg gggtagtcgg cgggtcaccc cgtgacggg  
31561 ctgttccagg gggccttcgg gccggtggcg gtcgcccacc accggctcct caccgggtc  
31621 cccgacgggt ggcgggcggg ggacgccga gccgtacca tcggttcc caccggccac  
31681 tacgcgctgc acgacctggc cgggttgtag gccgggcagt ccgtgctggt ccacggccgc  
31741 gccggcgggg tggggatggc tgccgtcgcg ttggcccggtc gggccggggc ggaggtgtc

FIG. 7-46

54/70

31801 gccacggcca gcccggccaa acaccgacg ctgcgggcgc tcggcctcga cgacgaccac  
31861 atcgccctcgt cccgggagag cgggttcggt gagcgggttcg ccgcgcgctac cggggggcgg  
31921 ggcgtcgacg tggtcctgaa ctgctcacc ggcgacctgc tcgacgagtc cgcgcggctg  
31981 ctcgccgacg ggcgggtctt gtcgagatg ggcaagaccg acctgcggcc ggcgagacag  
32041 ttccggggcc ggtacgtccc gttcgacctg gccgaggccg gtcccgatcg gctcggcgag  
32101 atcctggagg aggtcgtcgg tctgctggcc gccggtgcc tcgaccggtt gccggtgtcg  
32161 gtgtgggagt tgtcggcgcc cccggccgcg ctaccccaca tgagccgggg ccgacacgtg  
32221 ggcaagctcg tctcaccca gcccgccccc gtgacccccg acggaacggt gctggtcacc  
32281 ggcgggaccg gcacctggg gcggctggtc gcccgccacc tggtagaccgg gacgggcgta  
32341 cccacctcc tggtggccag ccggcgcggt ccggcgggcc cctgcgacac cgccgaccgg  
32401 gccgacgtcg aaggcctcgg cgcgaccatc gagatcgtcg cctgcgctgac cggggtggtg  
32461 gaggcgctcg cggcgctgct cgaactcgatc cccgcggacc gtccgctgac cggggtggtg  
32521 cacaccgccg gggtcctggc cgacgggctg gtacacctca tcgacgggac cgccaccgat  
32581 caggtcctgc gggccaagg cgacgcggcg tggcacctgc acgacctgac ccgggacgcg  
32641 gacctgagct tcttcgtgct gttctcgtcg gcggcgtcgg tgctggccgg tcccgggcag  
32701 ggcgtgtacg cggcggccaa cggggtcctc aacgccctgg ccgggcaacg gcggggccctc  
32761 ggactgcccg cgaaggcgct cgggtggggc ctgtggggcg aggccagcga gatgaccagc  
32821 ggcctcgggtg accggatcgc ccgtaccggg gtccgccgcg tcgccgaccga gcgggcgctg  
32881 gccctgttcg acgcggctct gcgcagcggc ggggagggtgc tggtcccgct gtctgtcgac  
32941 aggtcggcgc tgcgccgggc cgagtacgtc cccgagggtg tcgcggggcg ggtccggtcc  
33001 acgccacggg ccgccaacag ggccgagacc ccgggccggg gcctgctcga ccgtctcgtc  
33061 ggtgcacccg agaccgatca ggtggccgcg ctggccgagc tgggtccgctc gcacgcggcg  
33121 gcggtcgccg gctacgactc ggccgaccag ctgcccgaac gcaaggcggt caaggacctc

FIG. 7-47

55/70

33181 ggggttcgact cgctggcggc ggtggagctg cgcaaccggc tcggcgtcac caccggcgtg  
33241 cggctgccc cgcgctggt gttcgaccac cgcacaccgc tggcgggtggc cgaacacctg  
33301 cggtcggagt tggtcgccga ctccgcgccg gacgtcgggg tcggtgcgcg cctcgacgac  
33361 ctggaacggg cgctcgacgc cctgcccgcg gtggcagagc ggcagaccgc acgcccgcgc  
33421 ctggaggcgc tgctgcgccg tggtcagagc cgacgacccc cgagaccga gccagtgcg  
33481 atcagtgacg acgccagtga cgcagagctg ttctcgatgc tcgacaggcg tctcggcggg  
33541 ggaggggacg tctaggtgac aggtcgattc cgcccgcgcg cagtggaccg tacccgcctg  
33601 acaggtccac cgggttcgcg tcgcctcca caccgacgg cgggggtatc cacggaaggg  
33661 atccgatgag cgagagcagc ggcatgaccg aggaccgcct ccggcgctat ctcaagcgca  
33721 ccgtcgccga actcgactcg gtgacaggtc ggctcgacga ggtcgagtac cgggcccgcg  
33781 aaccgatcgc cgtcgtcggc atggcctgcc ggttccccgg ggtgtgggac tcgccggagg  
33841 cgttctggga gttcatccgc gacggtggtg acgcgatcgc cgaggcgccc acggaccgtg  
33901 gctggccgcc ggcaaccgca cccgcctcg gtggtctcct cgcgagaccg ggcgcgttcg  
33961 acgccgcctt cttcggcatc tcaccccgcg aggcgtcgc gacggacccc cagcagcgcc  
34021 tgatgctgga gatctcctgg gaggcgttgg agcgtgcggg ttctgacccg tcgagccctg  
34081 gcggcagcgc cggtggcgtc ttcaccgggtg tcggtgcggt ggactacgga ccaggcccg  
34141 acgaggcacc cgaggaggtg ctggctacg tcggcatcgg caccgcctcc agcgtcgcct  
34201 ccggacgggt ggcgtaacc ctgggggttgg aggtccagc cgtcacccgtc gacaccgcct  
34261 gctcctccgg gctcaccgcg gtgcacctgg cgatggagtc gctgcgccgc gacgagtgc  
34321 ccctggtcct cgccggtggg gtcaccgtga tgagcagccc ggtgcgttc accgagttcc  
34381 gcagccaggg cgggttggcc gaggacggcc gctgcaaacc gttctcccgc gccgccgacg  
34441 gcttcgggct cgccgagggg gccgggggtcc tggtgctcca acggctgtcc gtcgcccggg  
34501 ccgagggccg gccggtgctg gccgtactgc gtggctcggc gatcaaccag gacggtgcc

FIG. 7-48

56/70

34561 gcaacgggct caccgcgccg agcggccccg cccagcggcg ggtgatcagg caggcgttgg  
34621 agcggggcgcg gctgcgtccc gtcgacgtgg actacgtgga gggccacggc accggcaccc  
34681 ggctggggcga tccgatcgag gcgcacgccc tgctcgacac gtacggtgcc gaccgggaac  
34741 ccggccgccc gctctgggtc ggtatcggta agtccaacat cggtcacacc caggcggcgg  
34801 cgggggtggc cggggtgatg aagaccgtgc tggcgctgcg gcatcgggag atcccggcga  
34861 cgttgcaact cgacgagccc cggtggggg ccgactggga ccggggcggg gtgtcggtg  
34921 tgtccgagac ccggccctgg ccggtgggg agcggcccg cggggcggg gtgtcctcgt  
34981 tcggcatcag cggcaccaac gcgcacgtca tcgtcgagga ggcccgagc ccgcaggcgg  
35041 ccgacctga ccgaccccc gagggcgtcg cactgggttt ctccgcgcgc gacgagcggg  
35101 ccaccgccga gcccgggtcg cggctcgccg accgtctcac cgacgaccg gcccctcgt  
35161 ccctgcgcgc cagggcgcc cgtgtcacc ctggtgccac ctgggagcat cgggcggtcg  
35221 tgccgcgacac cgccttcacc ggtgtcacc gctctcgccg gcctccggg cgtcgcggg  
35281 tcgtcggcgg gggcgaggag ggcctcgccg gctccggg cgtcgcggg ggacgtccc  
35341 tcgacggagc cgtcagcggg cgggcgcgcg cggcccgccg ggtggtgctg gtcttcccc  
35401 ggcaaggcgc acagtggcag ggcattggcc ggacactgct gcggcagtcg ccgacctcg  
35461 cggagtccat cgacgcctgc gagcgggccc tcgccccgca cgtggactgg tcgctgcgcg  
35521 aggtgctcga cggcgagcag tcgttggacc ccgtcgacgt ggtgcagccg gtgctgttcg  
35581 cggtgatggt gtcgttggcg cggttgtggc agtcgtacgg ggtgactccg ggtgcggtg  
35641 tgggtcactc gcagggggag atcgccgccc cgacgtggc tggcgcttg tcgttggccg  
35701 acgccgccag ggtggtggcg ttgcgcagcc gggtgctgcg ccgtctcggg ggtcacggcg  
35761 gcatggcgtc gttcgggctc caccgcgacc aggcgcgcca gcggatcgcg cgcttcgcgg  
35821 gtgcgctgac tgtcgcctcg gtcaacgggt cccgttcggg ggtgctggcc ggggagaacg  
35881 gcccgttgga cgagctgac gccgagtgcg aggcggaggg cgtgaccgcc cgtcggatcc

FIG. 7-49



57/70

35941 ccgtcgacta cgcctcacac tccccgcagg tggagtcgct gcgtgaggag ctgctcgccg  
36001 cactggccgg ggtccgtccg gtgtcggccg ggatccccct gtactcgacc ctgaccggtc  
36061 aggtcatcga aacggcgacg atggacgccc actactggtt cgccaacctc cgggagccgg  
36121 tgcgcttcca ggacgccacc aggcagctcg ccgagggcggg gttcgacgcc ttctgtcgagg  
36181 tcagcccgca cccggtgttg acagtcggtg tcgagggccac cctcgaggca gtgctgcccc  
36241 ccgacgcgga tccgtgtgtc acaggcaccc tgcgccgcga acgcggcggc ctcgcgcagt  
36301 tccacaccgc gctcgccgag gcgtacaccc ggggggtgga ggtcgactgg cgtaccgcag  
36361 tgggtgaggg acgcccggtc gacctgcccgg tctacccgtt ccaacgacag aacttctggc  
36421 tcccgggtccc cctgggcccgg gtccccgaca ccggcgacga gtggcggttac cagctcgccct  
36481 ggcaccccgt cgacctcggg cggtcctccc tggccggacg ggtcctggtg gtgaccggag  
36541 cggcagttacc cccggcctgg acggacgtgg tccgcgacgg cctggaacag cgcggggcga  
36601 ccgtcgtgtt gtgcaccgcg cagtcgcgcg cccggatcgg ccccgccactc gacgccgtcg  
36661 acggcacccg cctgtccact gtggtctctc tgctcgcgt cgcgagggc ggtgctgtcg  
36721 acgaccccag cctggacacc ctcgcgttgg tccaggcgt cggcgacgc gggatcgacg  
36781 tccccctgtg gctggtgacc agggacgccc ccgccgtgac cgtcggagac gacgtcgatc  
36841 cggcccaggc catggtcggc gggctcggcc ggggtggtggg cgtggagtcc ccgcccgggt  
36901 ggggtggcct ggtggacctg cgcgagggcg acgccgactc ggcccgggtcg ctggccgcca  
36961 tactggccga cccgcgcggc gaggagcagt tcgcgatccg gcccgacggc gtcaccgtcg  
37021 cccgtctcgt cccggcaccc gcccgcgcgg cgggtacccg gtggacgccc cgcgggacccg  
37081 tcctggtcac cggcggcacc ggcggcatcg gcgcgcacct ggcccgtgg ctcgccggtg  
37141 cgggcgccga gcacctggtg ctgctcaaca ggcggggagc ggaggcggcc ggtgccgccc  
37201 acctgcgtga cgaactggtc gcgctcggca cgggagtcac catcacggcc tgcgacgtcg  
37261 ccgaccgcga ccggttggcg gccgtcctcg acgccgcacg ggcgcaggga cgggtggtca

FIG. 7-50



58/70

37321 cggcgggtgtt ccacgccgcc gggatctccc ggtccacagc ggtacaggag ctgaccgaga  
37381 gcgagttcac cgagatcacc gacgcgaagg tgcggggtac ggcgaacctg gccgaactct  
37441 gtcccagact ggacgccctc ctacgcggcg ggcaacgcct tcctcgacgc cgtggtcggc  
37501 ggctggcctc gcagtgggct gccggtcacc ctgatacgctt ggggtctgtg aacatggccg  
37561 gtaccgaggg cggcgactac accctggacg cgggggacct agggcctgag cgcagcggg  
37621 cgatcgagga gctgcggacc gactgttca gaaactgttca cggccggccc ctcttcgacg  
37741 tggaccggga ggtccgcgcc ggtccgcgcc ggggcccagg agaccggtca ggaatcggat ctgcccggc  
37801 aactcgggtg ggctggcgtc gatgccggag gccgaacgtc acgagcatgt cggccggctg gtccgagccg  
37861 aggtggcagc ggtgctgggc ggtgctgggc cacggcacgc cgacgggtgat cgagcgtgac gtcgccttcc  
37921 gtgacctggg attcgactcc ggtggccacg gtcggtgagc cgagggcgac gaaccggctc gggcgggtga  
38041 ccgggggtccg ggtggccacg ggtggccacg gtcggtgagc cgagggcgac gacagtggac cgcctcacccg  
38101 cgcactacct ggaacgactc ggaacgactc gtcggtgagc cgagggcgac gacccggct cgtcggcgcc  
38161 tcccgcaggc acccggggag ggtggccacg gtcggtgagc gccgacgagc gtcggtgagc gctgcccgc  
38221 tcgccgggtgg agtgcgtagc ggtggccacg gtcggtgagc gccgacgagc cccgaccagt tgtgggactt catcgtcgcc gacggcgacg  
38281 cggtcaccga gatgccgtcg gacccgtcct gacccgtcct gggacctcga cgcgctgttc gacccggacc  
38341 ccgagcggca cggcaccagc tactcccggc acggcgcggtt cctggacggg gggccgact  
38401 tcgacgcggc gttcttcggg atctcgccgc gtagggcggtt ggcgatggat ccgacgacg  
38461 ggcagggtcct ggagacgacg tgggagctgt tcgagaacgc cggcatcgac ccgactccc  
38521 tgcgcgggtac ggacacccgt gtcttcctcg gcgctgcgta ccaggggtac ggcagaacg  
38581 cgcagggtgcc gaaggagagt gaggggtacc tgctcacccg tggttcctcg gcggtcgcct  
38641 ccggtcggat cgcgtacgtg ttgggggttg aggggccggc gatcactgtg gacacggcgt

FIG. 7-51

59/70

38701 gttcgtcgtc gcttgtggcg ttgcacgtgg cggccgggtc gctgcgatcg ggtgactgtg  
38761 ggctcgcggt ggcggtggg gtgtcggtag tggccggtcc ggaggtgttc accgagttct  
38821 ccaggcaggg cgcgctggcc ccgacggtc ggtgcaagcc ctctccgac caggccgacg  
38881 ggttcggatt ggttcggggt cgcgagggc gtcgctgtgg tgctcctgca gcggttgtcg gtggcgggtg  
38941 gggaggggcg gtaatgggtt ggcggcgccg tcgggtgggtg ggtgaatcag gatggggcga  
39001 gtaatgggtt ggcggcgccg tcgggtgggtg cgcagcagcg ggtgattcgg cgggcgtggg  
39061 gtcgtgcggg gtcgtgcggg ggggatgtgg gtgtgggtga ggcgatggg acggggacgc  
39121 ggttggggga ggttgggggt tccggtggag ttggggggcgt aggcgaatgt gtatgggggtg ggtcgggggtg  
39181 ggggtgggtcc ggggtgggtg ggttcggtag ggttcggtag ggttcattgtg caggcggcgg  
39241 cgggtgtggt cgggtgtgac aaggtgggtg tgggggttggg tgggggggtg gtgggtccga  
39301 tgggtgtgtcg ggggtgggtg ccgggtgggtg ttgattgggtc gtcgggtggg ttgggtgggtg  
39361 cggatgggtt gcgggggtgg ccgggtgggtg tggatgggggt gcgtcgggggt ggggtgttcgg  
39421 cgtttggggt gtcgggggac aatgctcatg tgggtgggtgg gtaggcggccg gggtcgggtg  
39481 tggggggcga acggccggtg gagggggtcgt gagggggtt cgcgggggtg gctgggtgggtg  
39541 tgggtgccgt ggtgctgtcg gcaaaagacc aaaccgccct gaccgagctc gcccgacgac  
39601 tgcacgacgc cgtcgacgac accgtcgccc tcccggcggt gcccgccacc ctccgccaccg  
39661 gacgcgccca cctgccctac cgggccgccc tgctggcccg cgaccacgac gaactgcgcg  
39721 acaggctgcg ggcgttcacc actgggtcgg cggctcccgg tgtgggtgtcg ggggtggcgt  
39781 cgggtggtgg tgtgggtgtt gtcgggttccg gtcagggtgg gtatgggtg gggatggcgc  
39841 ggggggttgtt gtcgggttccg gtgttttgtgg agtcgggtgg agagtgtgat gcggtgggtg  
39901 cgtcgggtgtt ggggttttcg gtgttggggg tgttggaggg tcggtcgggt gcgctcgt  
39961 tggatcgggt ggatgtggtg cagccggtgt tgttcgtggt gatgggtgtcg ttggcggcgt  
40021 tgtggcggtg gtgtgggggt gtcgctgcgg cgggtgggtgg tcattcgcag ggggagatcg

FIG. 7-52

60/70

40081 cggcggcgggt ggtggcgggg gtgtgtgtcgg tgggtgatgg tgcgcgggtg gtggcgttgc  
40141 gggcgcgggc gttgcgggcg ttggccggcc acggcgggcat ggtctccctc gcggtctccg  
40201 ccgaacgcgc ccgggagctg atcgcacctt ggtccgaccg gatctcggtg gcggcgggtca  
40261 actccccgac ctcggtggtg gtctcggtg accacaggc cctcgccgc ctcgtcgccc  
40321 actgcgccga gaccggtgag cgggccaaga cgctgcctgt ggactacgcc tccactccg  
40381 cccacgtcga acagatccgc gacacgatcc gacacgatcc tcaccgacct ggccgacgtc acggcgcgcc  
40441 gacccgacgt cgccctctac cgccctctac tccacgctgc acggcgcccg gggcgccggc acggacatgg  
40501 acgcccggta ctggtacgac ctggtacgac aacctgcgt aacctgcgt cctcgacgag gccgtcgagg  
40561 ccgcccgtcg cgacgggtac cggtcttctg cggtcttctg tcgagatgag cccacacccg gtctctaccg  
40621 ccgcggtgca ggagatcgac ggagatcgac gacgagacgg gacgagacgg ctgctgcac cgggacacccg  
40681 gcgagcggca cctggtcgcc gaactcgcc gaactcgcc ttcccttgc gaactacccg ttcgaggcga  
40741 ggcgggcgat cctccccgcc cctccccgcc acacaccccg gactacccg gactacccg taccgcgtcg  
40801 cccggtactg gctcgccccg gctcgccccg acggcgcccg accaggtcgc accaggtcgc gtcttcggcg  
40861 actggcggcc cctgggccac cctgggccac cctgggcccg agctgtccgg cagctacctc gtcttcggcg  
40921 acgccccgga gaccctcggc gaccctcggc cacagcgtcg agaaggccgg cgggctcctc gtcccgggtg  
40981 ccgctcccga ccgggagtc ccgggagtc ctcgcggtcg ctcgcggtcg ccctggacga ggcggccgga cgcctcgccg  
41041 gtgtgctctc cttcgccgcc cttcgccgcc gacacccgca cccacctggc cgggacccga ctctcgccg  
41101 aggccgacgt cgaggcccca ctctggctgg ctctggctgg tcaccagcgg cggcgtcgca ctcgacgacc  
41161 acgacccgat cgactgcgac cgactgcgac caggcaatgg tgtgggggat cggacgggtg atgggtcttg  
41221 agaccccgca ccggtggggc ccggtggggc ggcctggtgg acgtgacct acgtgacct cgaacccacc gccgaggacg  
41281 ggggtggtctt cgccgccctc cgccgccctc ctggccgccg ggcctggtgg acgtgacct acgtgacct gccgaggacg  
41341 acggcatccg ccacgggccga cgcgtcgtcc cggctcgtcc ggcctcgtcc ggcctcgtcc ggcctcgtcc aacgccaggt  
41401 ggacaccggc gggcacggcg ctcgtcacgg gcgtacggg tgcctcggc tgcctcggc ggcacgtcg

FIG. 7-53

61/70

41461 cgcggtacct ggcccggtcc ggggtgaccg atctcgtcct gctcagcagg agcggccccg  
41521 acgcacccgg tgccgccgaa ctggccgccg aactggccga cctcggggcc gagccgagag  
41581 tcgagggcgtg cgacgtcacc gacggggccac gcctgcgcgc cctgggtgcag gagctacggg  
41641 aacaggaccg gccggtccgg atcgtcgtcc acaccgcagg ggtgcccagac tcccgcccc  
41701 tcgaccggat cgacgaactg gagtccgtca gcgccgcgaa ggtgaccggg gcgcgggctgc  
41761 tcgacgagct ctgcccggac gccgacacct tcgtcctgtt ctcctcgggg gcgggagtggt  
41821 ggggtagcgc gaacctgggc gcgtacgcgg cagccaacgc ctacctggac gccctggccc  
41881 accgccgccg ccaggcgggc cgggccgcga cctcgggtcg cctggggggcg tgggccggcg  
41941 acggcatggc caccggcgac ctcgacgggc tgacccggcg tgacctgcgg gcgatggcac  
42001 cggaccgggc gctgcgcgcc tgcaccaggc gttggaccac ccacgacacc tgtgtgtcgg  
42061 tagccgacgt cgactgggac cgcttcgccg tgggtttcac cgccgccccg ccagaccccc  
42121 tgatcgacga actcgtcac actcgtcac tccgcgccgg tggccgcccc caccgctgcg gcggcccccg  
42181 tcccggcgat gaccgccgac cagctactcc agttcacgcg ctcgcacgtg gccgcgatcc  
42241 tcggtcacca ggaccgggac gcggtcgggt tggaccagcc cttcacccag ctgggcttcg  
42301 actcgctcac cgccgtcggc ctgcgcaacc agctccagca ggccaccggg cggacgctgc  
42361 ccgccgccct ggtgttccag caccaccagg tacgcagact cgccgaccac ctcgcgcagc  
42421 agctcgacgt cggcaccgcc ccggtcgaag cgacggggcag cgtcctgcgg gacggctacc  
42481 ggcggggccgg gcagaccggc gacgtccggt cgtacctgga cctgctggcg aacctgtcgg  
42541 agttccggga gcgggttcacc gacgcggcga gcctggggcg gcctgaggaa ctcgtcgacc  
42601 tggccgacgg atccggcccc gtcactgtga tctgttgcgc tctgttgcgc ggcactgcg gcgctctccg  
42661 ggccgcacga gttcgcccga ctgcctcgg cgctgcgcgg caccgtgccg gtgcgcgccc  
42721 tcgcgcaacc cgggtacgag gcgggtgaac cggtgccggc gtcgatggag gcagtgcctg  
42781 ggggtgcaggc ggacgcggtc ctcgcggcac agggcgacac gccgttcgtg ctgggtcggac

FIG. 7-54



62/70

42841 actcggcgggg ggcctgatg gcgtacgcc gcgtacgccc tggcgaccga gctggccgac cggggccacc  
42901 cgccacgtgg cgtcgtgctc ctgcacgtgt accaccccg accaccccg tcaccaggag gcggtgcacg  
42961 cctggctcgg cagctgacc cccgccctgt gcgtacgaca ggctgaccgg caggtggcgt ccgagggaca  
43021 cccggctcac ccggtctgcc cacgtgggtg gtggccgcca gcgagccgat gggggagtgg ccgagcagc  
43081 gttggcagtc cgtggcagtc cacgtggccg ttcgggcacg gacgcgacat cgacgcctgg ttgagcgggg  
43201 cgatggtgca ggagcacgcc aacacgaccg atcgcccggt gctgggcccga cccgtaccga tgatccgggg  
43261 agagggcatg actgtactgg ggttacggca accgggggct accacgccac cgccgtgcgg gtgctcgacg accgacctt  
43321 actgtactgg cgacgacccg gtgtgaccg acccgccagg acgtgcacgc cgtgcctgct gcggtccgga cggcctcgac  
43381 cgacgacccg cgagacgtgg acccgccagg acccgccagg acgtgcacgc cgtgcctgct gcggtccgga cggcctcgac  
43441 cgagacgtgg acccgccagg acccgccagg acgtgcacgc cgtgcctgct gcggtccgga cggcctcgac  
43501 caccggggcc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc  
43561 cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc cgtggcgagc  
43621 gcaggaggtg ggtccgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac  
43681 ggtccgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac  
43741 gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac gctgcgcgac  
43801 ggcggtgacc caccgcccgc caccgcccgc caccgcccgc caccgcccgc caccgcccgc caccgcccgc  
43861 caccgcccgc caccgcccgc caccgcccgc caccgcccgc caccgcccgc caccgcccgc caccgcccgc  
43921 ggcggtggcg ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc  
43981 ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc ggtgctgcgc  
44041 ggtggtggcg ggtggtggcg ggtggtggcg ggtggtggcg ggtggtggcg ggtggtggcg ggtggtggcg  
44101 ccgtgacgcg ggtgacgcg ggtgacgcg ggtgacgcg ggtgacgcg ggtgacgcg ggtgacgcg  
44161 ccggggccctg tccggccagg cgggtcacc cgggtcacc cgggtcacc cgggtcacc cgggtcacc

FIG. 7-55

63/70

44221 gaccaccgcc gcactgcgca gcgtcgccaa ggcgctgccc ggtctaccg ccggtggccc  
44281 ggtcgtcagg cgacgtcgtt caccggtcct caccgaccacc gccactgcc cggtcgaact  
44341 ctgagggtgcc tgcgatgcgc gtcgtcttct cctccatggc cagcaagagc cacctgttcg  
44401 gtctcgttcc cctcgcctgg ccttccgcg gccttccgga cgggtacgg gtcgtcgcct  
44461 caccggctct caccgacgac atcacggcgg ccgactgac ggcgtaccg gtcggcacccg  
44521 acgtcgacct tgtcgacttc atgacccacg ccgggtacga catcatcgac tacgtccgca  
44581 gcctggactt cagcgagcgg gaccggcca cctccacctg ggaccacctg ctcggcacatgc  
44641 agaccgtcct caccgccgacc ttctacgccc tgatgagccc ggactcgctg gtcgagggca  
44701 tgatctcctt ctgtcggctg tggcgacccg actggctctc tggaccgcag accttcgccg  
44761 cgtcgatcgc ggcgacgggtg accggcgtgg ccacgcccgc actcctgtgg ggaccgcaca  
44821 tcacgggtacg ggcccggcag aagttcctcg ggctgctgcc cggacagccc gccgcccacc  
44881 gggaggaccc cctcgcgcgag tggctcacct ggtctgtgga gaggctcggc ggccgggtgc  
44941 cgaggagct cgaggagctg gtggtcgggc agtgagcat agtggacgat cgaccccgcc ccggtcggga  
45001 tgcgcctcga caccgggctg aggacgggtg gcatgcgcta cgtcgactac aacggcccct  
45061 cggtgggtgcc ggactggctg cagcagagc cgacccgccc aggggtctgc ctaccctgg  
45121 gcatctccag ccgggagAAC agcatcgggc aggtctcctt cgacgacctg ttgggtgcgc  
45181 tcggtgacgt cgacgccgag atcatcgcga cagtggacga gcagcagctc gaaggcgtcg  
45241 ccacgtccc ggccaacatc cgtacggctg ggttcgtccc gatgcacga ctgctgccga  
45301 cctgcgcggc gacggtgcac cacggcggtc ccggcagctg gcacaccgcc gccatccacg  
45361 gcgtgccgca ggtgattcctg cccgacggct gggacacccg ggtccgcgc cagcggacccg  
45421 aggaccaggg ggcggggcatc gccctgccgg tgcccagact gacctccgac cagctccgcg  
45481 aggcgggtcg gcgggtcctg gacgatcccg ccttcaccgc cggtcggcg cggatgcggg  
45541 ccgacatgct cgccgagccg tccccgccg aggtcgtcga cgtctgtgcg gggctggctcg

FIG. 7-56



64/70

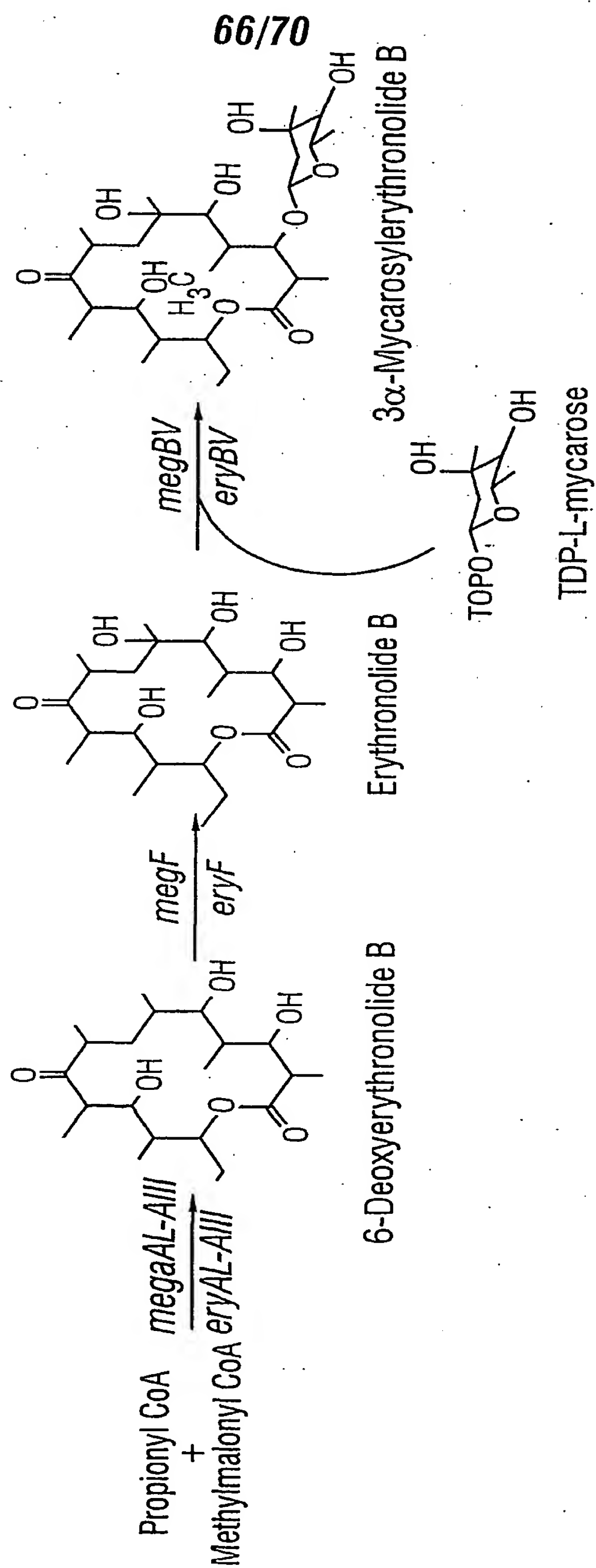
45601 gggaacggac cgccgtcggg tgagcaccga cgccaccac gtccggctcg gccggtgcgc  
45661 cctgctgacc agccggctct ggctgggtac ggcagccctc gccggccagg acgacgccga  
45721 cgcagtacgc ctgctcgacc ctgctcgacc acgcccgctc ccggggcgctc aactgcctcg acaccgccga  
45781 cgacgactct gcgtcgacca gcgtcgacca gtgccagggt cgccgaggag tcggtcggcc ggtggttggc  
45841 cggggacacc ggtcggcggg ggtcggcggg aggagaccgt cctgtcggtg acggtgggtg tcccaccggg  
45901 cgggcaggtc cgtgcggcgt ggcggggggc gcctctccgc ccggcagatc atcgccctct gtgagggtc  
45961 cctgcggcgt ctcggtgtcg accacgtcga cgtccttcac ctgccccggg tggaccgggt  
46021 ggagccgtgg gacgaggtct gacgaggtct ggcagccctc gtggccggcg gaaaggtctg  
46081 ttacgtcggg tcgtcgggct tcccggatg gcacatcgtc gccggccagg agcacgccgt  
46141 ccgccgtcac cgcctcggcc tgggtgtcca ccagtgtcgg tacgacctga cgtcgcgcca  
46201 tcccgaactg gaggctcctg ccgccgcga ggcgtacggg ctccgggtct tcgccaggcc  
46261 gacccgcctc ggcggtctgc tcggcggcga cggtcgggc cggcagccg gccgggcgtc  
46321 gggacagccg acggcactgc gctcggcggt gtagggcgta cgggtgttct gcagagacct  
46381 cggcgagcac cccgccgagg tcgcaactgg cgcactgctg gtgggtgctg tcccggcccg gtgtggcggg  
46441 ggcggtcgtc ggtgcgcgga cgcgggacg gctcgactcc gcgtcccg cctgcggcgt  
46501 cgccctcggc gcgacggaac tcaccgccct tcaccgccct cctgacctgc gggaacccgt  
46561 agggggcggc ccgagggcgt ggctacggtg agagcccgcc ccgcccggt cagccggtgg ggtgagccg  
46621 gtcggtgcgg cgggacggcc gccgcgggcc actcggccac ctcgccgacg tggtcggcga ggtagaagt  
46681 cagcaggccc aaggtccggg tacggccggg gactaccgag gactaccgag tacggcagcc agcgttgggc  
46741 cccgcccggg gtcgtcaacg ggtcgggtgt accgcagagg accgcagagg gtggtgatgc cggcccgag  
46801 gtcctccacc gctggtcaacg ggtcgggtgt accgcagagg accgcagagg gtggtgatgc cggcccgag  
46861 cggcgggccc gcctgccagg cgtaggagcg cagcaccggg tggtcggccc gcagcacccg  
46921 cagcgacatg tccaacagcc cctgggtcggc caatgcggcc tcgctgacct cgagcctgcg

FIG. 7-57

65/70

46981 catctgctcg acgagtcctgt cctcgtcggg caggtcgggtg cgccgctcgt ggaccggggg  
47041 ggcggtctgc ccggagacga acaaccgcag cggtcgcacc ccgggacgag cctccaggcg  
47101 acgggcggtc tcgtaggcga ccagggcgc ccagctgtga catgctgtga ccgaacaggg cgaacggaac  
47161 ctcgccgacg aggtcgcga gcacggccgc gacctcgtcg gacctctcc cggcggtgcc  
47221 gagagcccgc tcgtcacgtc ggtcctgccg gcccggtac tgcaccgcc acacgtcgac  
47281 ctccggggcc agtgcccggg cgaggtcga gtagagtcg gcggcggtc ccgcgtgcgg  
47341 gaagcagtac agccggggcc ggtgtccgc gcgggaccg aaccgccga accaggtgtt  
47401 catcggtgtc tcattccgtc ggtcgcaccg gcaggtgtc gatgccgc agcaggagcg  
47461 accgccgcca gacaacctc gccgagggga agcccagca cagcttcggg aagcggtcga  
47521 acagggcccc cagggcgacc tctccctcca gcttgccag cgggcggccc atgcagtagt  
47581 ggatgccgtg ccgaaagggt aggtgtcccc ggctgtccct ggtgacgtcg aaccggtcgg  
47641 ggtcggggaa ctgtccccgg ctgcgggttg ccgcccgtt gcgatacagg acggtgctgt  
47701 acgccgggat cgtcaccccc cgatctcca cctcggcggt ggcgaaccgg gtggtggtct  
47761 ccggtggggc ctggtagcgc aggatctcct ccaccgctc gggcagcagt gccgggtcct  
47821 tccggaccag cgcgagctgg tcgggggtgg tcagcagcag gtaggtgccg atcccgatga  
47881 ggctcaccga cgcctcgaat ccgcccagca gcagcaccag cgcgatggat gtgagttcgt  
47941 cgcggctgag ccggtcggcg tcgtcgtcct ggaccggat c  
(SEQ ID NO: 1)

FIG. 7-58



**FIG. 8A**

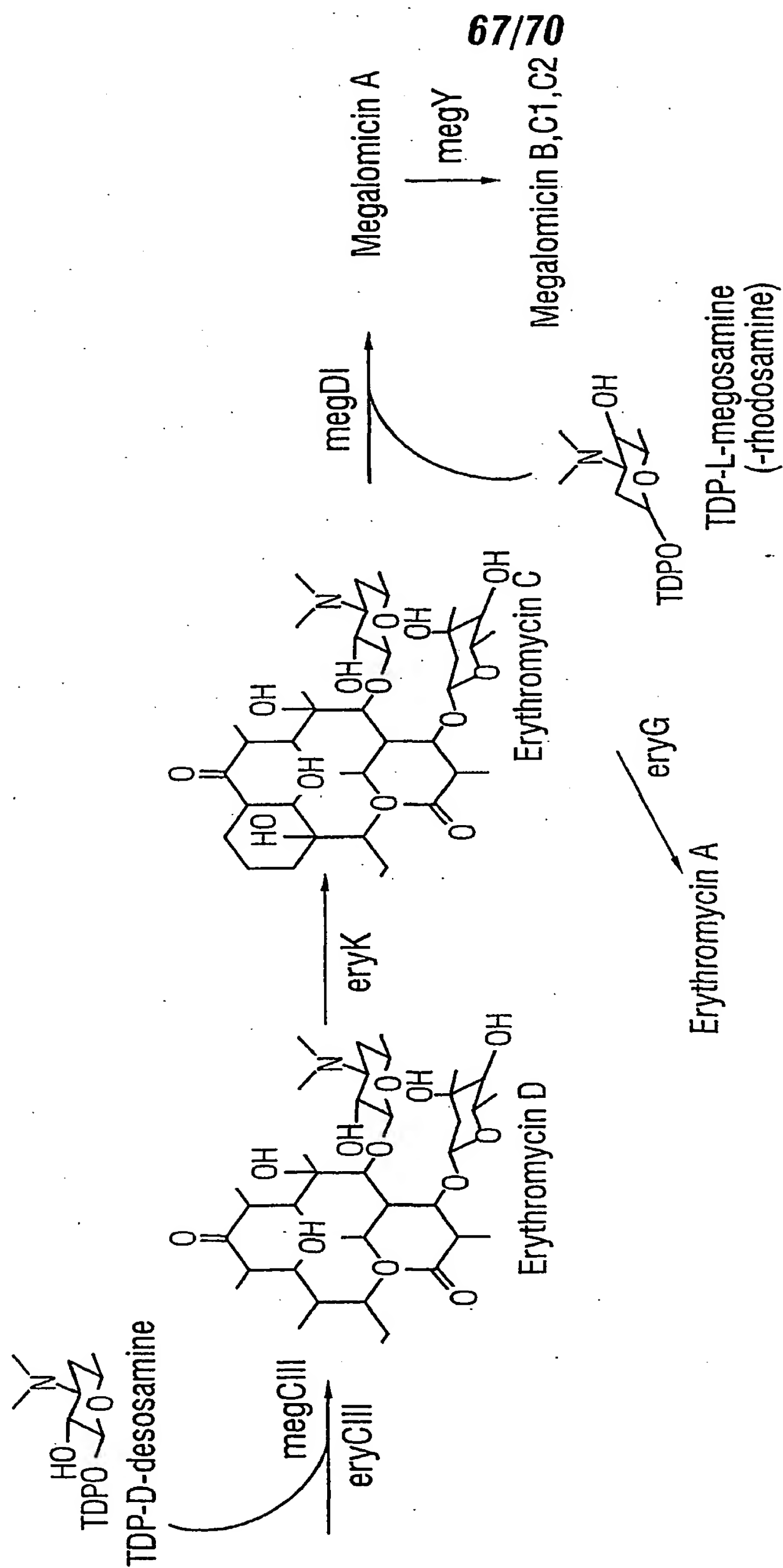
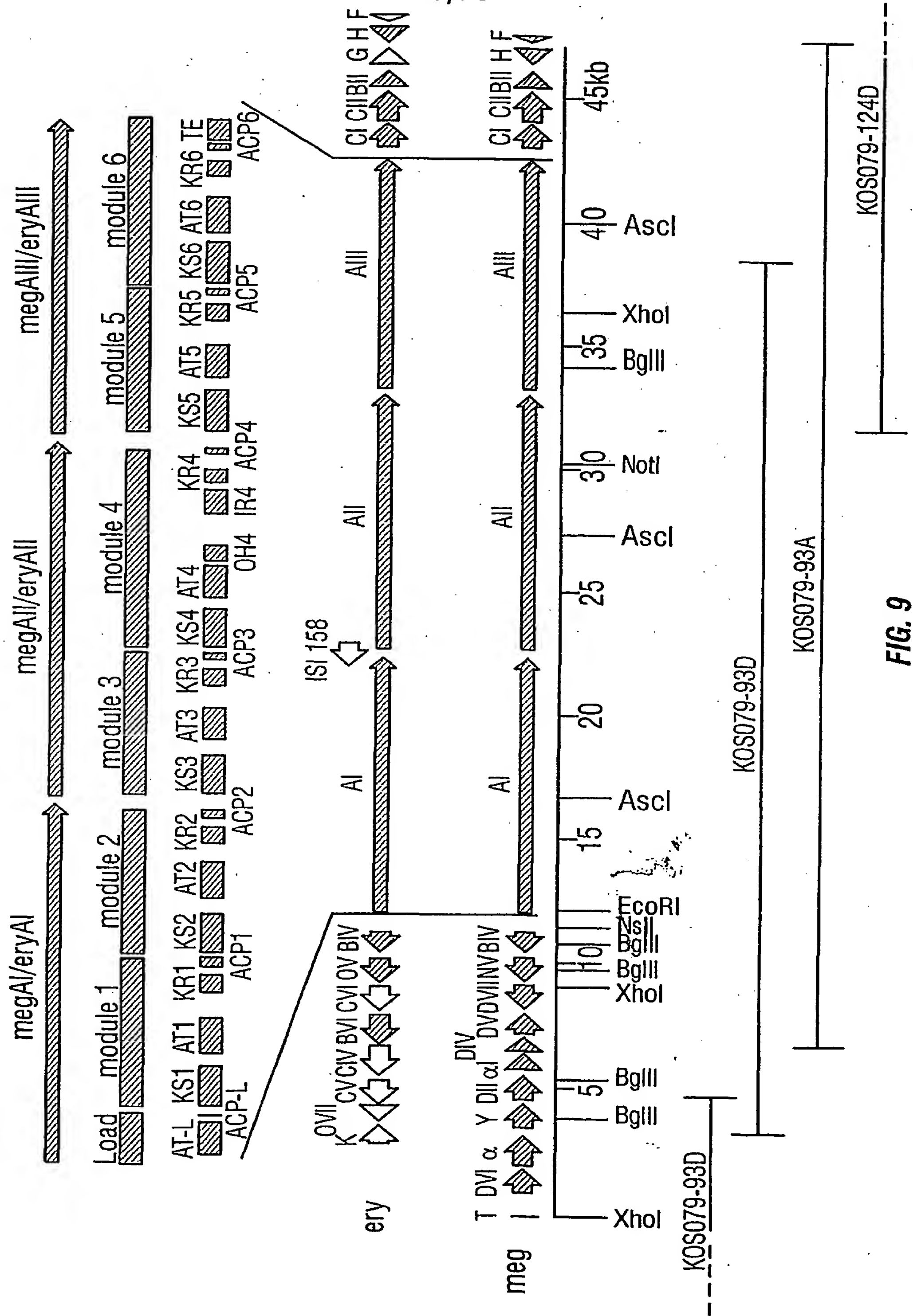


FIG. 8B



**FIG. 9**

69/70

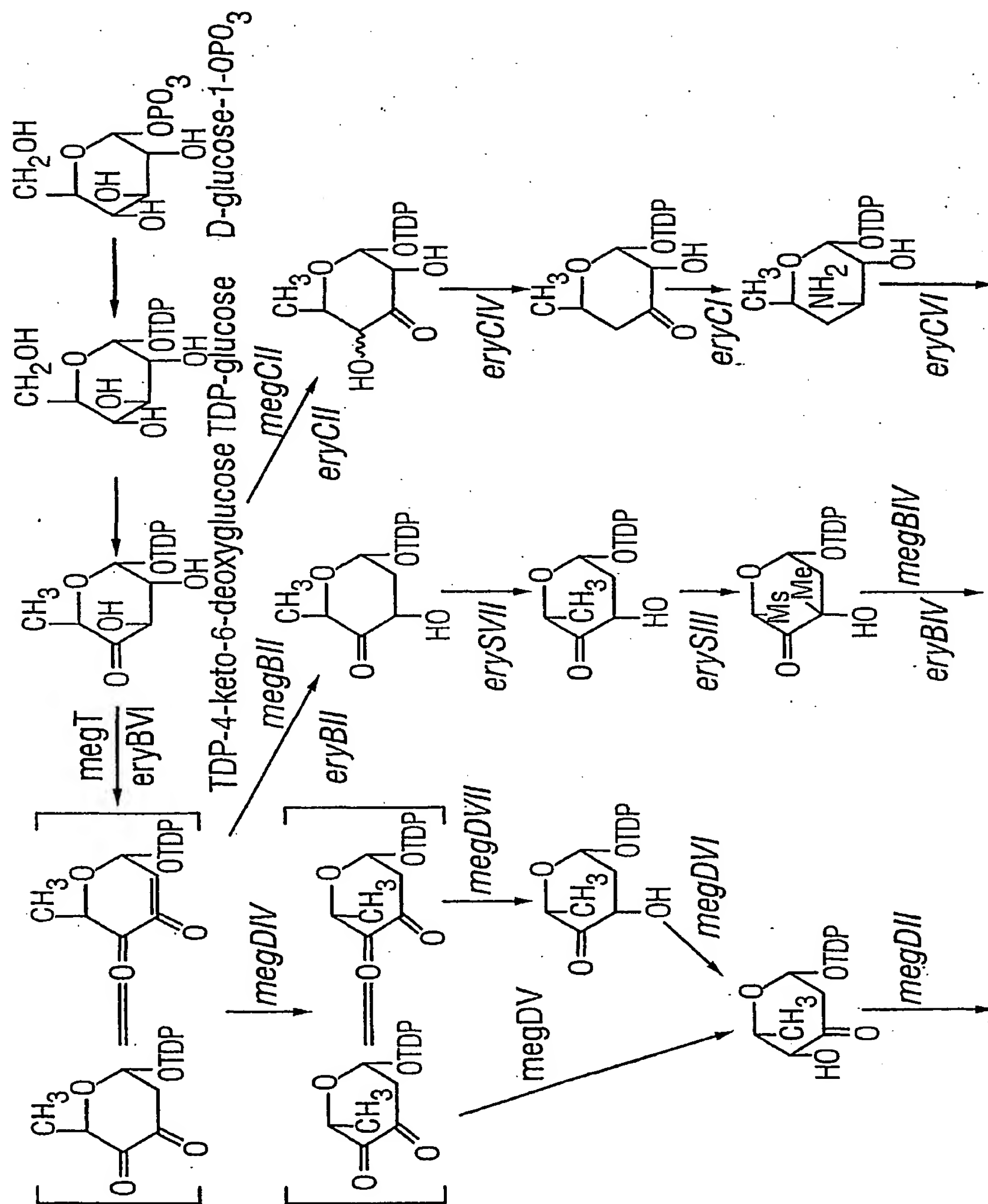


FIG. 10A



70/70

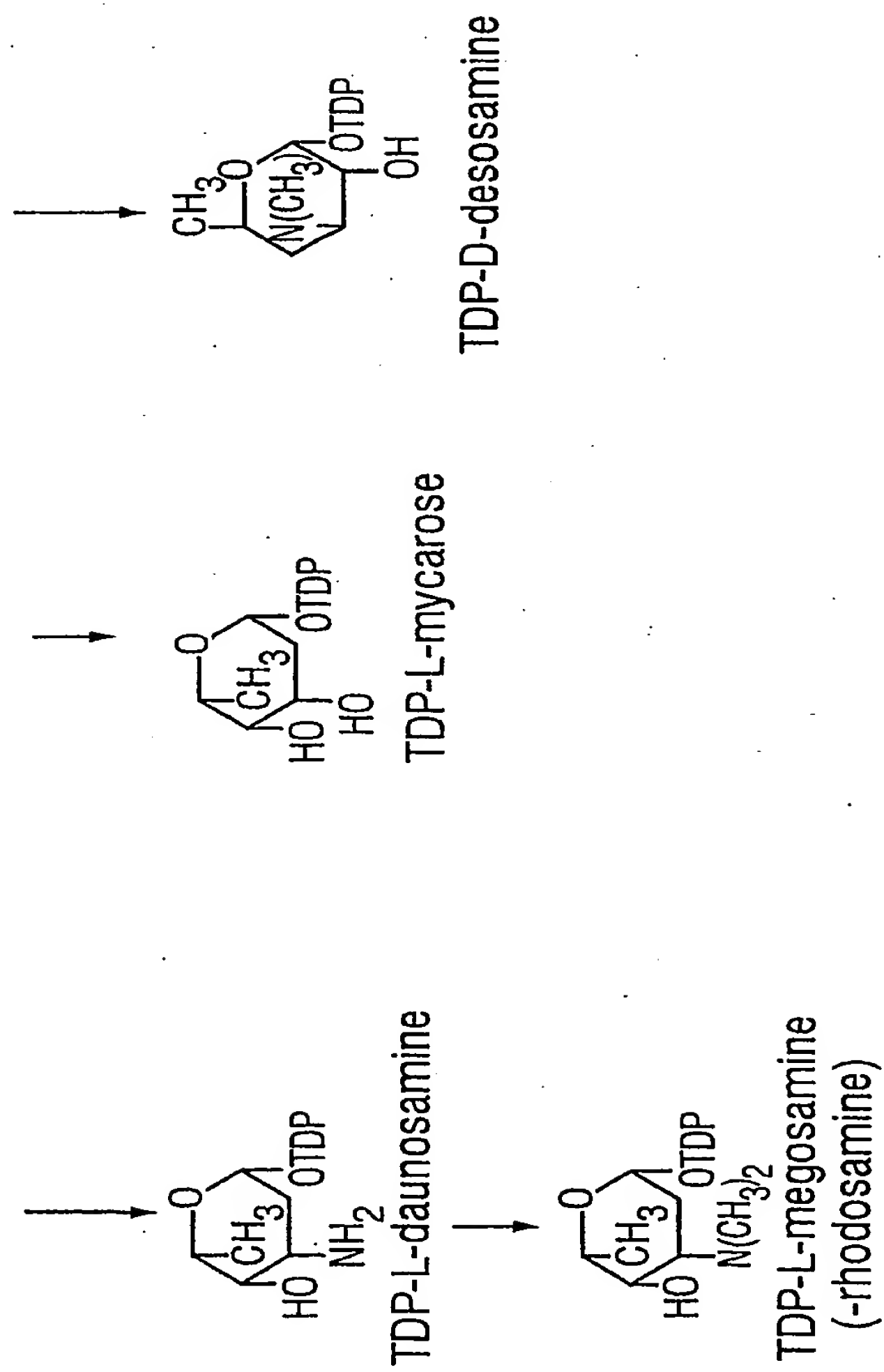


FIG. 10B

## SEQUENCE LISTING

<110> Kosan Biosciences, Inc.

<120> Recombinant Megalomycin Biosynthetic  
Genes and Uses Thereof

<130> 300622004740

<140> To be assigned  
<141> Herewith

<150> US 60/158,305  
<151> 1999-10-08

<150> US 60/190,024  
<151> 2000-03-17

<160> 34

<170> FastSEQ for Windows Version 4.0

<210> 1  
<211> 47981  
<212> DNA  
<213> Micromonospora megalomicea

<220>  
<221> CDS  
<222> (1)...(144)  
<223> megBVI(megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;  
SEQ ID NO: 2= translated amino acid sequence

<221> CDS  
<222> (928)...(2061)  
<223> megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase,  
TDP-4-keto-6-deoxyhexose 3,4-isomerase;  
SEQ ID NO: 3= translated amino acid sequence

<221> CDS  
<222> (2072)...(3382)  
<223> megDI, rhodosaminyl transferase (eryCIII homolog),  
TDP-megosamine glycosyltransferase;  
SEQ ID NO: 4= translated amino acid sequence

<221> CDS  
<222> (3462)...(4634)  
<223> megG(megY), mycarosyl acyltransferase, mycarose O-acyltransferase;  
SEQ ID NO: 5= translated amino acid sequence

<221> CDS  
<222> (4651)...(5775)  
<223> megDII, deoxysugar transaminase (eryCI, DnrJ homolog),  
TDP-3-keto-6-deoxyhexose 3-aminotransaminase;  
SEQ ID NO: 6= translated amino acid sequence

<221> CDS  
<222> (5822)...(6595)  
<223> megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);  
SEQ ID NO: 7= translated amino acid sequence

<221> CDS  
<222> (6592)...(7197)  
<223> megDIV, TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVII, dnmU homolog); TDP-4-keto-6-deoxyhexose 3,5-epimerase;  
SEQ ID NO: 8= translated amino acid sequence

<221> CDS  
<222> (7220)...(8206)  
<223> megDV, TDP-hexose 4-ketoreductase (eryBIV, dnmV homolog),  
TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 9= translated amino acid sequence

<221> CDS  
<222> (8228)...(9220)  
<223> megBII-1(megDVII), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;  
SEQ ID NO: 10= translated amino acid sequence

<221> CDS  
<222> (9226)...(10479)  
<223> megBV, mycarosyl transferase, mycarose glycosyltransferase;  
SEQ ID NO: 11= translated amino acid sequence

<221> CDS  
<222> (10483)...(11424)  
<223> megBIV, TDP-hexose 4-keotoreductase,  
TDP-4-keto-6-deoxyhexose 4-ketoreductase;  
SEQ ID NO: 12= translated amino acid sequence

<221> CDS  
<222> (12181)...(22821)  
<223> megAI; SEQ ID NO: 13= translated amino acid sequence

<221> misc\_feature  
<222> (12505)...(13470)  
<223> megAI, AT-L

<221> misc\_feature  
<222> (13576)...(13791)  
<223> megAI, ACP-L

<221> misc\_feature  
<222> (13849)...(15126)  
<223> megAI, KS1

<221> misc\_feature  
<222> (15427)...(16476)  
<223> megAI, AT1

<221> misc\_feature  
<222> (17155)...(17694)  
<223> megAI, KR1

<221> misc\_feature  
<222> (17947)...(18207)  
<223> megAI, ACP1

<221> misc\_feature  
<222> (18268)...(19548)  
<223> megAI, KS2

<221> misc\_feature

<222> (19876)...(20910)

<223> megAI, AT2

<221> misc\_feature

<222> (21517)...(22053)

<223> megAI, KR2

<221> misc\_feature

<222> (22318)...(22575)

<223> megAI, ACP2

<221> CDS

<222> (22867)...(33555)

<223> megAII; SEQ ID NO: 14= translated amino acid sequence

<221> misc\_feature

<222> (22957)...(24237)

<223> megAII, KS3

<221> misc\_feature

<222> (24544)...(25581)

<223> megAII, AT3

<221> misc\_feature

<222> (26230)...(26733)

<223> megAII, KR3 (inactive)

<221> misc\_feature

<222> (26998)...(27258)

<223> megAII, ACP3

<221> misc\_feature

<222> (27393)...(28590)

<223> megAII, KS4

<221> misc\_feature

<222> (28897)...(29931)

<223> megAII, AT4

<221> misc\_feature

<222> (29953)...(30477)

<223> megAII, DH4

<221> misc\_feature

<222> (31396)...(32244)

<223> megAII, ER4

<221> misc\_feature

<222> (32257)...(32799)

<223> megAII, KR4

<221> misc\_feature

<222> (33052)...(33312)

<223> megAII, ACP4

<221> CDS

<222> (33666)...(43271)

<223> megAIII; SEQ ID NO: 15= translated amino acid sequence

<221> misc\_feature

<222> (33780)...(35027)

<223> megAIII, KS5

<221> misc\_feature

<222> (35385)...(36419)

<223> megAIII, AT5

<221> misc\_feature

<222> (37068)...(37604)

<223> megAIII, KR5

<221> misc\_feature

<222> (37860)...(38120)

<223> megAIII, ACP5

<221> misc\_feature

<222> (38187)...(39470)

<223> megAIII, KS6

<221> misc\_feature

<222> (39795)...(40811)

<223> megAIII, AT6

<221> misc\_feature

<222> (41406)...(41936)

<223> megAIII, KR6

<221> misc\_feature

<222> (42168)...(42425)

<223> megAIII, ACP6

<221> misc\_feature

<222> (42585)...(43271)

<223> megAIII, TE

<221> CDS

<222> (43268)...(44344)

<223> megCII, TDP-4-keto-6-deoxyglucose 3,4-isomerase;  
SEQ ID NO: 16= translated amino acid sequence

<221> CDS

<222> (44355)...(45623)

<223> megCIII, desosaminyl transferase, desosamine glycosyltransferase;  
SEQ ID NO: 17= translated amino acid sequence

<221> CDS

<222> (45620)...(46591)

<223> megBII-2(megBII), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,  
TDP-4-keto-6-deoxyglucose 2,3 dehydratase;  
SEQ ID NO: 18= translated amino acid sequence

<221> CDS

<222> (46660)...(47403)

<223> megH, TEII; SEQ ID NO: 19= translated amino acid sequence

<221> CDS

<222> (47411)...(47980)

<223> megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid sequence

<400> 1

ctcgagecga tgctcggcgg cgcggtgggc caaccagtcg tggacgtcgt cgggtggcgg 60  
gggaggtccg ccgtgccgag tcaggaaacg tattgccgat tgtgtggatt ccggagtcgc 120

atgaccgttg	acccgatccc	ccatacgcct	ctcccgtgat	gtcgtgggcg	gtccgtgcgg	180
taccgcccgg	actgacattc	gtcgatcaag	accccgccca	gtgtagggct	ccgcccgcga	240
cgggagaagg	tccgtcgaa	aacttccggg	tgaccggtcg	ccggcgtcgg	tgaaacgggc	300
gtcggagcac	ccgatcattg	ctgtcgggtga	acttcctaac	tgtcggcgcg	cacatctttc	360
tgaccgggtg	gttccgtggg	atgacgcggt	cccggcccgt	ctggaactgt	gcgtgggact	420
gaccgggttg	ggcgtgtttt	cgcccgtttc	cgaactgcgg	attcgtcgat	cgcgaggtg	480
ggagcgggtg	gctgaccggg	atgatctgca	atcatggcgc	tcaatgacga	tctctttag	540
catgggtccg	gccgaggggt	cgacaggccc	gaaacgcccg	gcattccagc	tggttcgacga	600
cgtcgacatc	accgtgcaag	ccgcgatgac	accgacacca	cgccatgctg	gtgccgcact	660
ggaagggtgg	cgcgatcagg	gaaatggccg	tgctactaga	cagacgcca	acagctgtcc	720
gggcctgcgg	aaacagcatc	gatctgcgtc	agccgttcat	tgccccggcg	gcaccgcctt	780
ggaaatccgt	gccaccgggt	gtccgcagtg	acgatcgcg	acccgggttt	cgagacagca	840
ggtagtaggc	gatgcaggcg	tttcgtctcg	cgccggacgc	gtcgactag	gtggaatccg	900
tcacagtctt	caatccggga	gcgttctatg	gcagttggcg	atcgaaggcg	gctgggcccg	960
gagttgcaga	tggtccgggg	tctctactgg	gggttcgggtg	ccaacggcga	tctgtactcg	1020
atgctcctgt	ccggacggga	cgacgacccc	tggacctggg	acgaacgggt	gcgggcccgc	1080
ggacggggac	cgtacgccag	tcgggcccga	acgtgggtgg	tcggtgacca	ccggaccgcc	1140
gccgaggtgc	tcgccgatcc	gggcttcacc	cacggcccgc	ccgacgctgc	ccggtggatg	1200
caggtggccc	actgcccggc	ggcctcctgg	gccggcccct	tcggggagtt	ctacgcccgc	1260
accgaggacg	cggcgtcggt	gacagtggac	gccgactggc	tcagcagcg	gtgcgccagg	1320
ctggtgaccg	agctggggtc	gcgcttcgat	ctcgtgaacg	acttcgcccg	ggaggtcccg	1380
gtgctggcgc	tcggtaccgc	gcccgcactc	aaggggcgtg	accccgaccg	tctccgggtc	1440
tggacctcgg	cgacccgggt	atgcctggac	gcccagggtca	gcccgcaca	gctcgcggtg	1500
accgaacagg	cgctgaccgc	cctcgacgag	atcgacgcgg	tcaccggcgg	tcgggacgcc	1560
gcggtgctgg	tgggggtggg	ggcggagctg	gcggccaaca	cgggtgggcaa	cgccgtcctg	1620
gccgtcaccg	agcttcccga	actggcggca	cgacttgccg	acgacccgga	gaccgcgacc	1680
cgtgtggtga	cgaggtgtc	gcggacgagt	cccggcgctc	acctggaacg	ccgcaccgcc	1740
gcgtcggacc	gccgggtggg	cggggtcgac	gtcccagacc	gtggcgaggt	gacagtggtc	1800
gtcgccgcgg	cgaaccgtga	tcccagaggtc	ttcaccgatc	ccgaccggtt	cgacgtggac	1860
cgtggcggcg	acgccgagat	cctgtcgtcc	cgcccgggt	cgccccgcac	cgacctcgac	1920
gccctgggtg	ccaccctggc	cacggcgggc	ctgcggggcg	ccgcgcgggt	gttgccccgg	1980
ctgtcccgtt	ccgggcccgt	gatcagacga	cgtcgggtcac	ccgtcgcccg	tggtctcagc	2040
cgttgcccg	tcgagctgta	gaggaagaac	gatgcgcgtc	gtgttttcat	cgatggctgt	2100
caacagccat	ctgttcgggc	tgggtcccgt	cgcaagcgcc	ttccaggcgg	ccggacacga	2160
ggtacgggtc	gtcgccctcg	cggccctgac	cgacgacgtc	accggtgccg	gtctgaccgc	2220
cgtgcccgtc	ggtgacgacg	tggaaacttg	ggagtggcac	gcccacgcgg	gccaggacat	2280
cgtcgagtac	atgcggaccc	tcgactgggt	cgaccagagc	cacaccacca	tgctcctggga	2340
cgacctcctg	ggcatgcaga	ccaccttcac	cccagacctc	ttcgccctga	tgagccccga	2400
ctcgctcctc	gacgggatgg	tcgagttctg	ccgtcctgg	cgctccgact	ggatcgctctg	2460
ggagccgctg	accttcgccc	ccccgatcgc	ggcccgggtc	accggaaccc	cgcacgcccg	2520
gatgctgtgg	ggtccggacg	tcgccacccg	ggcccggcag	agcttccctg	gactgctggc	2580
ccaccaggag	gtggagcacc	gggaggatcc	gctggccgag	tggttcgact	ggacgctgcg	2640
gcgcttcggc	gacgaccgcg	acctgagctt	cgacgaggaa	ctggtgctgg	ggcagtggac	2700
cgtggacccc	atccccgagc	cgctgcggat	cgacaccggc	gtccggacgg	tgggcatgcg	2760
gtacgtcccc	tacaacggcc	cctcggtggg	gcccgcctgg	ctgttgccgg	aaccggaacg	2820
tcggcgggtc	tgccctgacc	tcggcgggtc	cagccgggaa	cacggcatcg	ggcaggtctc	2880
catcggcgag	atgttggaag	ccatcgccga	catcgacgcc	gagttcgtgg	ccaccttcga	2940
cgaccagcag	ttgggtcggcg	tgggcagcgt	tcgggcaaac	gtccgtaccg	ccgggttcgt	3000
gccgatgaac	gtcctgctgc	ccacctgcgc	ggccaccggt	caccacggcg	gcaccggcag	3060
ttggctgacc	gccgccatcc	acggcgtacc	gcagatcctc	ctctcggaag	ccgacaccga	3120
ggtgcacgcc	aagcagctcc	aggacctcgg	cgccggggctg	tcgctcccgg	tcgcggggat	3180
gaccgcccag	cacctgcgtg	gggcgatcga	gcgggttctc	gacgagccgg	cgtaccgcct	3240
cggtgccggag	cggatgcggg	acgggatgcg	gaccgacccc	tcgcccggcc	aggtgggtcg	3300
catctgtcag	gacctggccg	ccgaccgggc	ggcacgcggc	aggcagccgc	gtcgaaccgc	3360
cgagccgcac	ctgccgcgat	gacttcaccc	accaccggga	ccggctgatg	ccggtcccgg	3420
aatccacacg	ccgactttcc	ttctgacacg	agggggcccc	ggtgggttacc	tccaccaact	3480
tggacacgac	agcacggccg	gactgaact	cgttgaccgg	gatgcgggtc	gtcgccgcct	3540
tcctggtctt	cttcacgcac	gtcctgtcga	ggctcatccc	gaacagctac	gtgtacgcgg	3600
acggcctgga	cgccttctgg	cagaccaccg	gacgggtggg	ggtgtcgttc	ttctttatcc	3660
tcagcgggtt	cgtgctgacc	tggtcggcgc	gggccagcga	ctcgggtgtg	tcgttctggc	3720
gcagacgggt	ctgcaagctc	ttcccccaacc	acctgggtcac	cgccttcgcc	gcggtgggtg	3780



tgttcctggt	caccgggacg	gcggtgagcg	gtgaggcgct	gatcccgaac	ctcctgctga	3840
tccacgcctg	gttcccggcc	ctggagatct	ccttcggcat	caaccgggtg	agctggtcgt	3900
tggcctgcga	ggcggttctt	tacctgtgct	tcccgtgtgt	cctgttctgg	atctccggta	3960
tccgcccggg	gcggctgtgg	gcctggggccg	ccgtgggtgt	cgccgcgac	tgggcgggtac	4020
cggtggtcgc	cgacctcctg	ctgccgagtt	ccccgccgct	gatcccgggg	cttgagtact	4080
ccgccatcca	ggactgggtc	ctctacacct	tccctgcgac	gcggagcctg	gagttcatcc	4140
tccggatcat	cctggcccgc	atcctgatca	ccggtcgggtg	gatcaacgtc	gggctgctcc	4200
ccgcgggtgct	gttgttcccg	gtcttcttcg	tgcctcgtct	cttcctgcgg	gggtgtctacg	4260
ccatctctct	gtcgatgatg	atccttcccc	tgggttctgat	catcgccagc	ggcgcgacgg	4320
ccgacctcca	gcagaagcgc	accttcatgc	gtaaccgggt	gatgggtgtg	ctcggcgacg	4380
tctccttcgc	gctctacatg	gtccacttcc	tgggtgatcgt	ctacggggcg	gacctgctgg	4440
ggttcagcca	gaccgaggac	gccccgctgg	gtctcgcact	cttcctgatc	attccgttcc	4500
tccgggtctc	cctgggtgctg	tccgtggctgc	tgtacagggt	cgtcgagcta	cccgtcatgc	4560
gtaactgggc	ccgcccggcc	tccgcccggc	gcaaaccgcg	cacggaacc	gaacagacc	4620
cttcccggcg	gtaagaagga	cggtgcacgc	gtgaccacct	acgtctggtc	ctatctgttg	4680
gagtacgaga	gggaacgagc	cgacatcctc	gatgcgggtg	agaaggctct	cgccagtggc	4740
agcctgatcc	tccgtcagag	tgtggagaac	tccgagaccg	agtacgcccg	ctaccacggg	4800
atcgcgcact	gcgtgggcgt	cgacaacggc	accaacgctg	tgaactcgc	gctggagtcg	4860
gtaggtgtcg	gacgcgacga	cgaggtcgtc	acgggtctcca	acaccgcccg	ccccacagtc	4920
ctggccatcg	acgagatcgg	cgcccggccg	gtcttcgtgg	acgtccgcga	cgaggactac	4980
ctcatggaca	ccgacctggt	ggaggcggcg	gtcacccccgc	gtaccaaggc	catcgtcccg	5040
gtgcacctgt	acgggcagtg	cgtggacatg	acagccctgc	gggaactggc	cgaccggcgg	5100
ggcctcaagc	tccgtggagga	ctgcgccccg	gcccacgggtg	cccggcggga	cggtcggctg	5160
gcccgggacga	tgagcgacgc	ggcggccttc	tccgttctacc	cgacgaagg	cctcggcgcc	5220
tacggcgacg	gcggcgcggt	cgtcaccaac	gacgacgaga	cagcccgccg	cctgcgacgg	5280
ctgcgggtact	acgggatgga	ggaggtctac	tacgtcaccc	ggaccccggg	tcacaacagc	5340
cgccctcgacg	aggtgcaggc	cgagatcctg	cggcgcaaac	tgaccgggct	cgacgcgtac	5400
gtcgcgggtc	ggcgggcggt	cgcccagcgg	tacgtcgacg	ggctcggcca	cctccaagac	5460
tccgacggcc	tccgaactccc	agtgggtcacc	gacggcaacg	aacacgtctt	ctacgtgtac	5520
gtcgtccgcc	accgcgcgcc	cgacgagatc	atcaagcgtc	tccgggacgg	gtacgacatc	5580
tccctgaaca	tcagctaccc	ctggccgggtg	cacaccatga	ccggcttcgc	ccacctcggt	5640
gtcgcgtcgg	ggtcgctgcc	ggtcacccgaa	cggtggcccg	gcgagatctt	ctcccttccc	5700
atgtaccctt	ccctccctca	cgacctgcag	gacagggtga	tccgaggcgt	gcgggaggtc	5760
atcacccggc	tgtgacgagc	ccgcgtgtcg	tcagcgaaga	cccactctgg	aagggccgggt	5820
catgccgaac	agccactcga	ccacgtcgag	caccgacgtc	gcccogtacg	agcgggcgga	5880
catctaccac	gacttctacc	acggccgtgg	caagggatac	cgtgccgaag	ccgacgcgct	5940
cgtggaggtc	gcccgcgaagc	acaccccaca	ggcggcgacc	ctgctggacg	tggcctgcgg	6000
gaccggatcc	cacctggtcg	agctggcgga	cagcttccgg	gaggtgggtg	gggtcgacct	6060
gtcggccgcc	atgctcgcca	ccgcgcgccg	caacgacccc	gggcgggaac	tgcaccaggg	6120
cgacatgcgc	gacttctccc	tccgaccgag	gttcgacgtc	gtcacctgca	tgttcagctc	6180
caccggttac	ctcgtcgacg	aggccgaact	ggaccgtgcc	gtggcggaac	tggccgggtca	6240
cctcgcgcct	ggcggcaccc	tccgtcgtgga	gcccctgggtg	ttcccggaga	cgttccggcc	6300
cggctgggtc	ggggccgacc	tgggtcaccag	cggtgaccgg	aggatctccc	ggatgtcgca	6360
caccgtcccc	gcgggtctgc	ccgaccgcac	cgcctcccgg	atgaccatcc	actacacgggt	6420
gggggtcaccg	gaggccggga	tccgagcactt	caccgaggtg	cacgtgatga	ccctgttcgc	6480
ccgcgcgcgc	tacgagcagg	ccttccagcg	ggcgggcctg	agctgctcgt	acgtcggcca	6540
cgacctgttc	tccgcccggc	ttttcgtcgg	ggtcgcgcgc	gagccggggc	gggtgagggtc	6600
gaggagctgg	gcacgagggg	ggtcttcacc	ttcacccccgc	agacgttcgc	cgacgagcgg	6660
gggggtgttcg	gcacggcgta	ccaggaggac	gtgttcgtgg	cggcgctcgg	ccgcccgtcg	6720
ttcccgggtg	cccagggtcag	caccaccccg	tcccggcggg	gtgtgggtccg	gggggtgcac	6780
ttcacgacga	tgcggggtc	catggcggaag	tacgtctact	gcgccagggg	tagggcgatg	6840
gacttcgccc	tccgacatccg	gcccgggttcc	ccgaccttcg	gcccggccga	gcccgtcgag	6900
ctctccgcgc	agtcgatgggt	cgggctgtac	cttcccgtgg	gcacgggcca	cctgttcgtc	6960
tccctggagg	acgacaccac	cctcgtctac	ctgatgtccg	ccggttacgt	ccccgacaag	7020
gaacggggcg	tgcaccccct	ggatccggag	ctggcggttc	cgatcccggc	cgacctcgac	7080
ctcgtcatgt	ccgagcggga	ccgggtcgca	cccaccctcc	gggaggcccc	ggaccagggg	7140
atcctgcccc	actacgcgc	ctgcggggcc	gcccgcgacc	gggtgggtcg	gacgtgaccc	7200
cggccggggcg	tgcggggccg	tgggtgggtg	cggcgcgctc	ggtttcctgg	gttcggcggt	7260
cacccacgcc	ctggccgacc	tcccgggtgcg	ggtgcggctc	gtcgcggggc	gggaggtcgt	7320
cgtgccctcc	ggtgccgtcg	ccgactacga	gacgcaccgg	gtggacctca	ccgaaccggg	7380
agcgctcgcg	gaggtgggtcg	cggacgcccc	ggcggtcttc	ccgttcggcg	cccagatcag	7440

gggtacgtca	gggtggcgga	tcagcgagga	cgacgtggtc	gccgaacgga	cgaaacgtcgg	7500
cctgggtccg	gacctgatcg	ccgtcctgtc	ccgctcgccg	cacgcccccg	tggtggtctt	7560
cccgggcagc	aacacgcagg	tcggcagggg	caccgcccgc	cgggtcatcg	acggcagcga	7620
gcaggaccac	cccagagggcg	tctacgacag	gcagaaacac	accgggggaa	agctgctcaa	7680
ggaggccact	gcggccgggg	cgatccgggc	gaccagtctg	cggctgcccc	cgggtgttcgg	7740
ggtgcccggc	gccggcaccg	ccgacgaccg	gggggtggtc	tccaccatga	tccgtcgggc	7800
cctgaccggc	caaccgctga	cgatgtggca	cgacggcacc	gtccggcggtg	aactgctgta	7860
cgtgaccgac	gccgcccggg	ccttcgtcac	cgccctggac	cacgcccagc	cgctcgccgg	7920
acgccacttc	ctgttgggga	cggggcggtc	ctggccgctg	ggcgaggtct	tccaggcggt	7980
ctcgcgacgc	gtcgcccggc	acaccggcga	ggaccgggtg	ccgggtgggt	cgggtgccgccc	8040
tccggcgcac	atggaccggt	cggacctgcg	cagcgtggag	gtcgaccccc	cccgggtcac	8100
ggctgtcacc	gggtggcggg	ccacgggtcac	gatggcgagg	gcggtcgacc	ggacgggtggc	8160
ggcgttgccc	ccccgcccgg	ccgcccggcc	gtccgagccc	tcctgaccgg	ggtcaccggg	8220
gttcgtccta	cggcaccggc	ccgtcgacgg	ccgggtgccg	gaagatcgct	tcgagttccc	8280
ggagttcctc	ctcgcccagc	gtcagctcgg	cggcccgtaa	cgccgagtcg	agctgctcgg	8340
gtgtgcgggg	gccgatgaca	gcgcccagga	tcccggggcg	ggacaggacc	caggccagac	8400
cgacctcggc	cgggtccgcg	ccgaggcgtc	ggcagtagtc	ctcgtacgcc	tcgacgaggg	8460
ggcgtacggc	ggggaggagc	acctggggcg	gtccctgcgc	cgacttgacg	gcgggttcgg	8520
ctgccaactt	ctccagtacg	ccgttgagca	gcccggcggtg	cagggggggac	caggcgaaca	8580
cgcccacccc	gtacgcctgg	gcggcgggca	ggacgtccag	ctcgggggtg	cggacggcca	8640
ggttgtagac	gacttggtgg	gagatcatgc	cgagcaggtt	gcggcggtgcc	gcgctctcct	8700
gggcgggcgg	gatgtgccag	cccggccaggt	tggaggagcc	gacgtacccc	accttcccac	8760
tgccgaccag	atgttcggcg	gcctgccaca	cctcgtccca	cgggtgcggcg	cggtcgatgt	8820
ggtgcgtctg	gtagatgtcg	atgtgggtcg	ccccgaggcg	gcggaggagg	ttctcgcagg	8880
cggcgacgat	gtgtcggggc	gagagcccgc	cgctcgttgac	ccgttcgctc	atctcgtgc	8940
ccaccttgg	cgccaggacg	gtctcctcgc	gtcgacctcc	gccctggggc	aaccaccgtc	9000
cgacgagttc	ctcgggtgtg	cccttgtaga	gccgccagcc	gtagatgtcg	gcgggtgtcg	9060
tgcagttgac	gccccgctcg	agggcggtgt	ccatcagccc	cagcgcgtcg	tcgtcgggtca	9120
cccgtccact	gaagttcacg	gtgccgagcc	agagtcggct	ggtgtgcaac	gccgatcgtc	9180
cgacgcgtac	ccgggcccgg	ccggcccccg	tggttcccac	gtcggtcacc	tgctcggcgcg	9240
gtgctggtgg	gcgagcgcc	ccagcacggg	tacgacctcg	gcggggggtcg	gcgcggccag	9300
cgcctcctgc	cgcagcttct	cggcgttctc	ggcgtgggaa	cggtcctcga	ccactgtggc	9360
gagagcctgc	cagaggggtg	cggcgctcgac	ctcgtccgga	cggaggaaga	cacccgctcc	9420
cagctcggcg	gtgcgctgac	cacgcaggac	acagtcacc	tcgtgggcca	cggagatctg	9480
cggtagcggc	tggtgcagcg	cgggtggcca	gcttccggca	ccgccgtggg	ggatgacggc	9540
ggcacagccc	ggcagcagga	tggtcatggg	aacgaagtc	accaggcgga	cgttggtccgg	9600
caccgacgcc	ggatcgagcc	cggagcgggt	caccacgac	tcgccgtcga	accgcgcgag	9660
ggtggccagt	gtccggagga	actcctgcgg	gttcgaggtg	atgccacgcg	ccgagtatcc	9720
cccgggtgaag	cagaccccgg	ggactccgtc	cgaggtcctg	agccactgcg	gcacgacgga	9780
ggacccgttg	tagggcaaag	tccgggtgtg	caccgactcc	agtcgggtct	ccaggcgga	9840
gctctcgggc	agctggtcga	cgctccactg	tccgacagcg	aggtcctcgc	tgtagtcgag	9900
gccgaaccgg	ccggcgacct	cgggtgagcca	gccgcccagc	gggtccggcc	ggtcgtcggc	9960
gggacgctgc	ccgcgcagg	cctgggagcg	gctgcggaag	tagccgggtg	ggtcgtcggc	10020
ccacagcagc	cgggcgtggg	cggccccgca	ggccttgggc	gcgaccgccc	cggcgaaggt	10080
gaagggtctc	cagagcacca	ggtcgggacg	ccagtcctatg	gcgaactcga	caggttcgtc	10140
gacgaaggag	tcgttggtga	ccaccgggaa	gacgaaccgg	gaggtggcct	cctcgatgcc	10200
gtgcagggaac	tcccacgagc	gcagttccgg	tccgcgtcgg	gcgaagtcca	ggtcgggtgg	10260
gtagcgggtg	acctgcgcgg	cggcctcagg	ggagatgtcg	aagagtcggg	ggtccgagcc	10320
gagtggcacc	gaggtcagtc	ccgcgcccag	gacgacgtcg	gtgagctcgg	gctgactggc	10380
cacccgagcg	tcgtggccgg	cgggtgtgag	cgcccaggcc	agggggagca	ggccctggaa	10440
gtgggtacgg	tgcgcgaaac	aggtgagcag	gacccgcaact	ggtcactcct	tggtcgagat	10500
gagggcgggc	acgggtccgg	cgatgccctc	ggccagcggc	acccgggggt	gccagccggg	10560
cagcgtccgg	aactcgggtg	agtcgaagtc	gtcgtcgcg	aagtcgttgg	cctcggcggt	10620
ctccgggtga	gggacgctga	cgacgggcac	cgcagggttg	ccgggtctgac	gtgccacgct	10680
ggcggcgacg	gtctcgaaga	tctcgccgag	gggtcggggc	tcgtccgcgc	tcggcggtcca	10740
gacgtcgccg	accagcgcc	cgtgggtgtg	cagtgcggcg	gtgaacgcgg	tggtccacgct	10800
ctcgacgtgc	aggaggttgc	ggcgcacgct	gccctcgtgc	cacatcgtga	tcggctcacc	10860
ggcgagggtc	cgcgggatca	tggcggtgac	gacaccccgg	ccgggtctgcc	ccgacggggc	10920
gctgtggccg	tagatcgcg	gcaggcgag	gatcaccggc	tcgacgaccc	cgctcctcgg	10980
ggcctgacgc	aggatccgct	cggcctcgat	cttgtgctgg	gcgtaccggc	tggggcgggc	11040
gggggttcg	gcctgggtgg	tgctggcgaa	caggagcacc	ggcgcggggtc	cgggtcttgc	11100

ccgcagcgcg	gcgacgaggt	cgcgcatgat	gcccgcgttg	acgcgttcgg	cctcggggcac	11160
cgtggcgggcg	ctgcgccagg	tcgacccgcc	ggcggcgtag	gcgaccagat	gcacgacgac	11220
gtcgggtgtcg	gcgacgacct	gcgcgacccg	gcccgggttcg	agcagggtcga	ctcgaagggtg	11280
ctcgatcccg	gcgctgcctg	gtggctggtc	gcgagacccg	gtgcgcgcga	cggcccgcag	11340
tcggagaggg	tgtgtggtaa	attcgcaag	aagggcgctt	ccgacgaatc	cagaaacgcc	11400
gagaagtgtg	acatgtcttg	tcattctacta	atgcattccg	atagccaccg	gcgcatggaa	11460
tccatttggt	ccccccagg	tgggtgcggg	tgacaaatcc	ggcctcagg	cggcctcaag	11520
cctcttttcga	gcgggtgctg	aggcttcccg	cgtaccctcg	gtggcctgcg	ttcggggcg	11580
tgtcggggaa	agggcggtac	gaggagtctg	gtagggcgctc	gcggcgcgta	ctccgggact	11640
gatccgggtc	gacgccccga	cgcgtgacag	ggcgctcgatc	cgtgccgccc	gtaccgccc	11700
ttttcggcga	tggtcgcaga	ttcctcccg	cgtgggtggac	tcattgggtc	tcccgggtgt	11760
ggccgcaccg	tcgggtggct	cgtcgggggt	gtcggagacc	gggtcgatcg	ccgtccccg	11820
ccgtgccgac	cagggtcgg	ccgtcgccga	ggtgggtcac	cgtcgggtgg	acccgggtccg	11880
ccggcgccca	ccgcccgatc	gtgcccacct	tcgcctccgc	gggtaaatagc	ttcgtcgatc	11940
tgatcgacac	ttccggcgac	gctatcaccg	gagcattccc	cggcaccacc	ggtcgatgcc	12000
tcgcgctttc	caaacaggg	aaacagcagc	tcacagcggt	tccaggcgcc	gggcaatcct	12060
agcgaagagt	ctcgatgggg	tcaaggtgaa	ttctgtcaca	gatgtttttg	ttaaatgtac	12120
tttcttcagc	caccctcgac	gttcatacaa	ttggccggca	tctctaccaa	gggggagtga	12180
gtggttgacg	tgcccgatct	actcggcacc	cggactccgc	acccaggggcc	gctcccattc	12240
ccgtggcccc	tgtgcgggtca	caacgaaccg	gagctgcggg	cccgcgcgcg	tcaattgcac	12300
gcatactctg	aaggcatttc	cgaggatgac	gtggtggccg	tcggcgccgc	cctcgcgcgc	12360
gagacacgcg	cgcaggacgg	gcccgaaccg	gccgtcgtcg	tggcctcctc	ggtcaccgag	12420
ctgaccgccc	cgctcgccgc	cctcgcccag	ggccgcccac	acccctcgg	ggtacgcggt	12480
gtcgcgccgac	ccacggcacc	ggtggtgttc	gtcctgccc	gtcagggcgc	ccagtggccc	12540
ggcatggcga	cccgaactgt	cgcgagtcg	cccgtcttcg	ccgcggcgat	gcgggcctgc	12600
gagcgggcct	tcgacgaggt	caccgaactg	tcgttgaccg	aggctcctgga	ctcacccgag	12660
cacctgcgc	gcgtcgaggt	ggtccagccc	gcgctcttcg	cgggtgcagac	ctcactggcc	12720
gccctgtggc	ggtcgttcgg	ggtgcgaccc	gacgcgcgtac	tcggacacag	catcggtgag	12780
ctggccgccc	ccgaggtctg	cggcgccgctc	gacgtcgagg	ccgcgcgcgc	ggccgcgcgc	12840
ctgtggagcc	gcgagatggt	cccactggtg	ggccgggggtg	acatggcggc	ggtggcgctc	12900
tccccggccg	agctggcagc	ccgggtcgag	cgggtgggacg	acgacgtcgt	gccggccggg	12960
gtcaacggtc	cccggtcgg	gctgctcacc	ggcgtctccc	agcccacgc	acggcggggtc	13020
gcccgaactg	cggcacagg	cgtacgcgc	caggtcgtca	acgtgtcgat	ggcggcgcac	13080
tcggcgccag	tcgacgcggt	cgcgagggc	atgcgctcgg	cgtgacctg	gttcgcccc	13140
ggcgactccg	acgtgcccta	ctacgcgggc	ctcaccggcg	ggcggtcgga	cacccgggaa	13200
ctcggcgccc	accactggcc	gcgcagtttc	cggctcccgg	tgcgcttcga	cgaggcgacc	13260
cgtgcgggtcc	tggaaactgca	gcccggcacg	ttcatcgagt	cgagcccgc	cccgggtgctg	13320
gcggcctccc	tgcagcagac	cctcgacgag	gtcgggtccc	cggccgcgat	cgtgccgacc	13380
ctgcaacgcg	accagggcgg	tctgcggcgg	ttcctgctcg	ccgtggcgca	ggcgtacacc	13440
ggtggcggtga	cagtcgactg	gaccgcgcgc	taccccgggg	tgacccccg	ccacctgccg	13500
tcggccgctcg	ccgtcgagac	cgacgaggg	ccctcgacgg	agttcgactg	ggccgcgcgc	13560
gaccacgtac	tgcgcgcgcg	gctgctggag	atcgtcggcg	ccgagacggc	cgcgctcgcc	13620
gggcgggagg	tcgacgccc	ggccaccttc	cgggaactgg	gcctcgactc	ggtcctcgcg	13680
gtgcagctgc	ggacccgcct	cgcacggcg	accgggcggg	atctgcacat	cgccatgctc	13740
tacgaccacc	cgaccccgc	cgcctcacc	gaggcgctgc	tgcgcggccc	gcaggaggag	13800
ccggggcggg	gtgaggagac	ggcacacccg	acggaggccg	aaccgcagca	acccgtcgcc	13860
gtggtcgcca	tggcgtgccg	gctgcccggc	ggcgtcacct	caccggagga	gttctgggag	13920
ctgctggccg	aggggcggga	cgcgctcggc	gggctgccc	ccgaccgggg	atgggacctg	13980
gactcgctgt	tccaccggga	cccgaaccgg	tcgggcacgg	cgcaccagcg	cgtggtggc	14040
ttcctcaccg	gcgcacacctc	cttcgacgct	gccttcttcg	ggctgtcgcc	acgggaggca	14100
ctggccgctcg	agccgcagca	gcggatcacg	ttggagctgt	cgtgggaggt	gctggaacgc	14160
gccgggatcc	ccccgacgtc	gttgccggacc	tcccggaccg	gggtgttcgt	cgtgctgatc	14220
ccccaggagt	acggcccccg	gctggccgag	gggggtgagg	gcgtcgagg	ctacctgatg	14280
accgggacca	ccaccagcgt	cgcctccggt	cgggtcgcc	acaccctcg	cctggagggg	14340
ccggcgatca	gcgtcgacac	cgcctgctcg	tcgtcgctcg	tcgcgctgca	cctggcgctgc	14400
cagtcgctgc	ggcgcggcga	gtcgacgatg	gcgctcgccg	gtggcggtgac	ggtgatgccg	14460
acaccgggca	tgctcggtga	cttcagtcgg	atgaactccc	tcgccccga	eggacgggtcc	14520
aaggcggttct	cggccgcgcg	cgaagggttc	ggcatggccg	aaggcgaggg	gatgctcctg	14580
ctggaacggc	tctcgagcgc	ccgcgcgcac	ggccaccccg	tgctcgccgt	gatcaggggc	14640
accgctgtca	actccgacgg	cgcgagcaac	ggactctccg	ccccgaacgg	ccgggcccag	14700
gtccgggtga	tccgacaggc	cctcgccgag	tccgggctga	cgcgccacac	cgtcgacgtc	14760



gtggagaccc	acggcaccgg	cacccgcctc	ggtgatccga	tcgaggcacg	ggcgtctctc	14820
gacgcgtacg	gcggtgaccg	tgagcaccgg	ctgcggatcg	gctcgggtcaa	gtccaacatc	14880
gggcacaccc	aggccgcccgc	cgggtgtcgc	ggtctgatca	aactggtgtt	ggcgatgcag	14940
gccggtgtcc	tgccccgcac	cctgcacgcc	gacgagccgt	caccggagat	cgactgggtcc	15000
tcgggcgcgga	tcagcctgct	ccaggagccc	gctgcctggc	ccgcccggcga	gcggccccgc	15060
cgggcccggg	tgctctcggt	cggcatcagc	ggcaccaacg	cacacgcgat	catcgaggag	15120
gcgcccgcga	ccggtgacga	cacccgaccc	gaccggatgg	gcccgggtgg	gccctgggtg	15180
ctctcggcga	gcaccggcga	ggcggtgcgc	gcccggggcg	cgcggtggc	cgggcaccta	15240
cgcgagcacc	ccgaccagga	cctggacgac	gtcgctact	cgctggccac	cggtcgggccc	15300
gcgctggcgt	accgtagtgg	gttcgtgccc	gccgacgcgt	ccacggcgct	gcggatcctc	15360
gacgaactcg	ccgcccgtgg	atccggggac	gcggtgaccg	gcaccgcccg	cgccccgcag	15420
cgcgctcgtct	tcgtcttccc	cggcccaggga	tggcagtggg	cggggatggc	agtcgacctg	15480
ctcgacggcg	acccggctct	cgcctcgggt	ctgcgggagt	gcgcccgcgc	gttggaaaccg	15540
tacctggact	tcgagatcgt	cccgttctct	cgggcccagg	cgacgcgcgc	gacccccgac	15600
cacacgctct	ccaccgaccg	cgtcgacgtg	gtccagccgg	tgctgttcgc	ggtgatggtg	15660
tccttgccgg	cccgggtggc	ggcgtagcgg	gtggaaccgg	cggccgtcat	cggacactcc	15720
cagggggaga	ttgccgcggc	gtgtgtggcc	ggggcgctct	cgctggacga	cgcgcccggg	15780
gcggtggccc	tgcgcagccg	ggtcatcgcc	accatgcccg	gcaacggcgc	gatggcctcg	15840
atcgccgcct	ccgtcgacga	ggtggcgccc	cggatcgacg	ggcgggtcga	gatcgccgcc	15900
gtcaacggtc	cgcgcgcggg	ggtggtctcc	ggcgaccgtg	acgacctgga	ccgcctggtc	15960
gcctcctgca	ccgtcgaggg	ggtgcggggc	aagcggctgc	cgggtggacta	cgcgtcgcac	16020
tcctcgacag	tcgagggcgt	ccgtgacgcg	ctccacgcgc	aactcggcga	gttcgggccc	16080
ctgcggggct	tcgtgccgtt	ctactcgaca	gtcaccggcc	gctgggtcga	gcccgcgcaa	16140
ctcgacgcgc	ggtactgggt	tcgcaacctg	cgccacaggg	tccggttcgc	cgacgcggtc	16200
cgtccctcg	ccgaccaggg	gtacacgacg	ttcctggagg	tcagcgccca	cccgggtgctc	16260
accacggcga	tcgaggagat	cggtgaggac	cgtggcggtg	acctcgtcgc	tgtccactcg	16320
ctgcgacgtg	gggcccggcg	tcccgtcgac	ttcggctccg	cgctggcccc	cgccttcgtg	16380
gccggcgctc	cagtggactg	ggagtccggc	taccagggtg	ccggggcgcg	tcgggtgccc	16440
ctgcccacgt	acccgttcca	gcgtgagcgc	ttctggttgg	aaccgaatcc	ggcccgcagg	16500
gtcgccgact	ccgacgaegt	ctcgctccctg	cggtagccga	tcgaatggca	cccgaaccgat	16560
ccgggtgagc	cgggacggct	cgacggcacc	tggctgctgg	cgacgtaccc	cggtcggggcc	16620
gacgaccggg	tcgaggcggc	gcggcaggcg	ctggagtccg	ccggggcgcg	ggtcgaggac	16680
ctggtggtgg	agccccggac	gggcccgggtc	gacctggtgc	ggcggtcga	cgccgtgggt	16740
ccggtggcgg	gcgtgctctg	cctgttcgct	gtcgcgagc	cggcggccga	acactccccg	16800
ctggcggtga	cgctcgttgc	ggacacgctc	gacctgaccc	aggcggtggc	cgggtcgggc	16860
cgggagtgtc	cgatctgggt	ggtcaccgag	aacgcgcgtc	ccgtcggggc	cttcgaacgg	16920
ctccgcgacc	cggcccacgg	cgcgctctgg	gccctcggtc	gggtcgtcgc	cctggagaac	16980
cccgcgcgtct	ggggcggcct	ggtcgacgtg	ccgtcgggtt	cggtcgccga	gctgtcgcgt	17040
cacctcggga	cgaccctgtc	cggcgcccggc	gaggaccagg	tcgccctccg	acccgacggg	17100
acgtacgccc	gccggtgggt	cagggcgggc	gcgggcggca	cgggcccgggtg	gcagccccgg	17160
ggcacggtgc	tcgtcaccgg	cggcaccggc	ggggtcggtc	ggcacgtcgc	ccggtggctg	17220
gcccgcaggg	gcaccccgtg	cctggtgctg	gccagccgcc	ggggaccgga	cgccgacggg	17280
gtcgaggagc	tactcaccga	actcgccgac	ctgggcaccc	gggccaccgt	caccgcctgc	17340
gacgtcaccg	accgggagca	gctccgtgcc	ctcctcgcga	ccgtcgacga	cgagcaccgc	17400
ctgtcggcgg	tggtccacgt	cgcgcgcgac	ctcgacgacg	gcaccgtcga	gacctcacc	17460
ggtgaccgca	tcgaacgggc	caaccggggc	aagggtgctc	gtgcccgcga	cctgcacgag	17520
ctgacccggg	acgccgacct	cgacgcgttc	gtgctcttct	cctcctccac	cgcgcggttc	17580
ggcgcgccgg	ggctcggcgg	ctacgtcccg	ggcaacgcct	acctcgacgg	tctcgcccag	17640
cagcgacgca	gcgagggact	cccggccacc	tcgggtggcgt	ggggtacctg	ggcgggcagc	17700
gggatggccg	agggtccggt	cgccgaccgg	ttccgcccgg	acggggtcat	ggagatgcac	17760
cccgaccagg	ccgtcgaggg	tctccgggtg	gcaactggtg	agggtgaggt	agccccgatc	17820
gtcgtcgaca	tcagggtggga	ccggttcctc	ctcgcgta	ccgcgcagcg	ccccaccggg	17880
ctcttcgaca	ccctcgacga	ggcccgtcgg	gccgcgcccg	gtcccgcgcg	cgggcccggg	17940
gtggcgggcg	tgggccgggt	gcccgtcggg	gaacgcgaga	aggcggtcct	cgacctggtg	18000
cggacgcacg	cggctgccgt	cctcggccac	gcctcggccg	agcaggtgcc	cgtcgacagg	18060
gccttcgccc	aactcggcgt	cgactcgcgt	tcggccctgg	aactgcgcaa	ccggctgacc	18120
actgcgacgg	gggtccggct	ggccacgacg	acggtcttcg	accacccgga	cgtacggacc	18180
ctggccggac	acctggccgc	cgaactgggc	ggcggtatcg	ggcgggagcg	gcccgggggc	18240
gagggcccga	cgggtggccc	gaccgacgag	ccgatcgcca	tcgtcgggat	ggcctgccgg	18300
ctgccggggg	gagtggaact	accggagcag	ctgtgggagt	tgatcgtctc	cgggcccggg	18360
accgcctcgg	cggcaccggg	ggaccgggag	tgggatccgg	cggagttgat	ggtctccgac	18420

acgacgggca	cccgtaccgc	cttcggcaac	ttcatgcccg	gggcgggcga	gttcgacgcg	18480
gcgttcttcg	ggatctcgcc	gcgtgaggcg	ttggcgatgg	atccgcagca	gcggcacgcc	18540
ctggagacca	cctgggaggg	gctggagaac	gccggtatcc	ggcccagatc	gttgcgcggt	18600
acggacaccg	gtgtcttcgt	gggcatgtcc	catcaggggt	acgccaccgg	ccgcccgaag	18660
cccgaggacg	aggtcgacgg	ctacctgttg	acaggcaaca	ccgcgagcgt	cgcctccggt	18720
cggatcgcg	acgtgttggg	gttggagggg	ccggcgatca	ctgtggacac	ggcgtgttcg	18780
tcgtcgcttg	tggcggttga	cgtggcgggc	ggttcggttg	gttctgggga	ctgtggtctg	18840
gcggtggcgg	gtgggggtgc	ggtgatggcc	ggtccggagg	tgttcaggga	gttctcccgg	18900
cagggcgcg	tggctccgga	cggcaggtgc	aagcccttct	cggacgaggg	cgacggcttc	18960
ggtctggggg	aggggtcggc	cttcgtcgtg	ttgcagcggt	tgtcgggtgg	ggtgcgggag	19020
gggcgtcggg	tgttgggtgt	ggtggtgggt	tcggcggtga	atcaggatgg	ggcgagtaat	19080
gggttggcgg	cgccgtcggg	ggtggcgag	cagcggtga	ttcggcgggc	gtgggggtcgt	19140
gcgggtgtgt	cgggtgggga	tgtgggtgtg	gtggaggcgc	atgggacggg	gacgcggttg	19200
ggggatccgg	tggagttggg	ggcgttgttg	gggacgtatg	gggtgggtcg	gggtggggtg	19260
ggtccggttg	tgggtgggttc	ggtgaaggcg	aatgtgggtc	atgtgcaggc	ggcgggcggt	19320
gtggtgggtg	tgatcaaggt	ggtgttgggg	ttgggtcggg	ggttgggtgg	tccgatggtg	19380
tgtcgggggt	ggttgtcggg	gttgggtgat	tggtcgtcgg	gtgggttgg	ggtggcggat	19440
ggggtgcggg	ggtggccggt	gggtgtggat	ggggtgcgtc	ggggtgggg	gtcggcggtt	19500
ggggtgtcgg	ggacgaatgc	tcattgtggtg	gtggcgagg	cgccgggggtc	ggtggtgggg	19560
gcggaacggc	cgggtggagg	gtcgtcgcgg	gggttgggtg	gggtgggtgg	tgggtgtggtg	19620
ccggtggtgc	tgtcggcaaa	gaccgaaacc	gccctgcacg	cccaggcacg	tcgactcgcc	19680
gaccacctgg	agacgcaccc	cgacgtcccg	atgaccgacg	tgggtgtggac	gctgacgcag	19740
gcccgcacac	gcttcgacag	gcgcgcggtc	ctcctcgccg	ccgaccggac	ccaggccgtg	19800
gaacggctgc	gcggcctcgc	cgggggcgaa	ccggggaccg	gtgtgggtgc	gggggtggcg	19860
tcgggtggtg	gtgtgggtgt	tgtttttcct	ggtcaggggtg	gtcagtgggt	ggggatggcg	19920
cgggggttgt	tgtcgggttc	ggtgttttgt	gagtcgggtg	tggagtgtga	tgcggtggtg	19980
tcgtcgggtg	tgggggtttc	ggtgttgggg	gtgttggagg	gtcggtcggg	tgcgccgtcg	20040
ttggatcggg	tggatgtggt	gcagccgggtg	ttgttcgtgg	tgatggtgtc	gttggcgcg	20100
ttgtggcggt	ggtgtgggg	tgtgcctgcg	gcgggtggtg	gtcattcgca	gggggagatc	20160
gcggcgcgcg	tgggtggcg	ggtgttgcg	gtgggtgatg	gtgcgcgggt	ggtggcggtg	20220
cgggcgcggg	cgttgcgggc	gttggccggc	cacggcgga	tggcctcgg	acgccgaggg	20280
cgcgacgacg	tacagaagct	cctcgacagc	ggccctgga	cggggaagct	ggagatcgcc	20340
gcggtcaacg	gcccgcagcg	ggtggtggtc	tccggcgacc	cccagaccgt	gaccgagctg	20400
gtcgagcact	gtgacgggat	cggggtccgg	gcccggacga	tccccgtcga	ctacgcctcc	20460
cactccgcac	aggtcgagtc	gctccgggag	gagctgctct	ccgtcctggc	cgggatcgag	20520
ggccgcccgg	cgacggtgcc	gttctactcc	accctcaccg	gtgggttcgt	cgacggcacc	20580
gaactggacg	ccgactactg	gtaccgcaac	ctgcgccacc	cgggtgcggt	ccacgcccgc	20640
gtcgaggcgc	tggcagcgcg	tgacctcacc	acgttcgtcg	aggtcagccc	gcaccccgtg	20700
ctgtcgatgg	cggtcgggga	gacgcttgcc	gacgtggagt	ccgcggtcac	tgtgggcacc	20760
ctggaacgcg	acaccgacga	cgtcgagcgc	ttcctcacct	ccctcgccga	ggcgcacgtc	20820
cacggcggtac	ccgtggactg	ggcggcggtc	ctcggctccg	gaaccctggt	cgacctgccc	20880
acctatccct	tccagggacg	gcggttctgg	ctgcaccccg	accgtggtcc	gcgtgacgat	20940
gtcgccgact	ggttccaccg	ggtcgactgg	acggcgacgg	ccaccgacgg	gtcggcccga	21000
ctcgacgggtc	gctgggtggt	ggtcgtaacc	gaggggtaca	cggacgacgg	ctgggtcgtg	21060
gaggtgcggg	ccgccctcgc	cgcgggtggt	gccgagccgg	tgggtgacgac	ggtcgaggag	21120
gtcaccgacc	gggtcgggtga	cagcgacgcg	gtgggtgtcga	tgtcggggct	ggccgacgac	21180
ggtgcggccg	agaccctggc	gctgctgcga	cgactcgacg	cacaggcgtc	caccaccca	21240
ctgtgggtgg	tcaccgtggg	ggcggtcgcc	cccgcgggtc	cgggtgcagcg	ccccgaacag	21300
gcgacggtgt	gggggttggc	ccttgtcgcc	tccctggaac	gcggacaccg	gtggaccggc	21360
ctgctggatc	tgccgcagac	accggacccg	cagctacgac	cccggctggt	cgaggcgctc	21420
gccggtgcgg	aggaccaggt	agcgggtccg	gccgacgccc	tacacgcccg	tcggatcgct	21480
cccaccccgg	tcaccggagc	cgggcccgtac	accgccccgg	gcgggacgat	cctcgtcacc	21540
gggggcaccg	ccggtctggg	tgcggtcacc	gcccgatggc	tcgccgagcg	cgggtgccga	21600
cacctcgccc	tggtcagccg	gcgcggggcg	ggcaccgccc	gcgtcgacga	ggtggtccgg	21660
gacctgaccg	ggctcggcgt	acgggtgtcg	gtgcactcct	gcgacgtcgg	cgaccgcgag	21720
tcggtcggcg	ccctggtgca	ggagttgaca	gcagccgggtg	acgtggtccg	gggggtggtc	21780
cacgtgcgg	gtctgcccc	gcaggtgcc	ctgaccgaca	tggacccggc	cgacctcgcc	21840
gacgtggtgg	ccgtgaaggt	cgacggcgcg	gtgcacctgg	ccgacctgtg	cccggaggcc	21900
gaactgttcc	tgctgttctc	ctcgggggccc	ggggtgtggg	gcagtgcccg	tcagggtgcg	21960
tacgccgccc	gaaacgcctt	cctggacgcc	ttcgcccagc	accggcgggg	ccggggtctg	22020
cccgccacct	cgggtggcgtg	ggggctctgg	gcggccgggg	ggatgacagg	ggaccaggag	22080

gcggtgtcgt	tcctgcgtga	gcggggcgta	cggccgatgt	cggtgccgag	ggcactggaa	22140
gcgctggaac	gggtcctcac	cgccggggag	accgcggtgg	tcgtcgccga	cgtcgactgg	22200
gcggccttcg	ccgagtcgta	cacctccgcc	cgcccccggc	cgctgctcca	ccggctcgtc	22260
acacctgcgg	cggcggtcgg	cgagcgcgac	gagccgcgtg	agcagaccct	ccgggaccgg	22320
ctggcgggccc	tgccccgggc	cgagcggtcg	gcggagctgg	tacgcctggt	ccggcgggac	22380
gccgcagccg	tgctcggcag	cgacgcgaag	gccgtacccg	ccaccacgcc	gttcaaggac	22440
ctcgggttcg	actcgtctgg	cgcggtccgg	ttccgtaacc	ggctggccgc	ccacaccggt	22500
ctgcgtctgc	cggccaccct	ggtcttcgag	caccggaacg	ccgcagccgt	cgccgacctc	22560
ctccacgacc	gactcggcga	ggccggcgag	ccgacccccg	tccggtcggt	gggcgcggga	22620
ctggccgcgc	tggagcaggc	cctgccccgac	gcctccgaca	cggagcgggt	cgagctggtc	22680
gagcgcttg	aacggatgct	cgccgggctc	cgccccgagg	ccggagccgg	ggccgacgcc	22740
ccgaccgccg	gtgacgacct	gggggaggcc	ggcgtcgacg	aactcctcga	cgcgctcgaa	22800
cggaactcg	acgccagggt	aaccggaact	gaccgcagcc	gcagccgaag	cagagaccga	22860
ggacctgtga	ctgacaacga	caagggtggc	gagtacctcc	gtcgtgcgac	gctcgacctg	22920
cgggccgccc	gcaagcgcc	gcgcgagctg	caatccgacc	cgatcgcggt	cgtcggcatg	22980
gcctgcggcc	taccggggcg	ggtgcacctc	ccgcagcacc	tgtgggacct	cctgcgccag	23040
gggcacgaga	cggtgtccac	cttccccacc	gggcgcggct	gggacctggc	cggtctcttc	23100
caccgagacc	ccgaccaccc	cggcaccagc	tacgtcgacc	ggggtgggtt	cctcgacgac	23160
gtggcgggct	tcgacgcgga	gttcttcggg	atctccccgc	gcgaggccac	ggccatggac	23220
ccgcaacagc	ggctgctgtt	ggagaccagt	tgggagctgg	tggagagcgc	cggcacgat	23280
ccgcaactcc	tgctgggcac	cccgaccggc	gtcttcctcg	gcgtggcgcg	gctcggtac	23340
ggcgagaacg	gcaccgaagc	cggtgacgcc	gagggctatt	cggtgaccgg	ggtggcacc	23400
gctgtgcct	ccgggcggat	ctcctacgcc	ctcgggctgg	agggctcgtc	gatcagcggt	23460
gacaccgcgt	gctcgtcgtc	gttggtggcg	ctgcacctgg	cggtcgagtc	gctgcggctg	23520
ggcgagtcga	gtctcgtgt	cgtcggcggg	gcggcggtca	tggcgacacc	aggggtgttc	23580
gtcgacttca	gccgccagcg	ggcggtggcc	gctgacggca	ggtcgaaggc	cttcggggcc	23640
gccgccgacg	ggttcggctt	ctccgagggg	gtctccctcg	tcctgctcga	acggctctcc	23700
gagggcgaaa	gcaacggcca	cgagggtgtg	gctgtcatcc	gtggctccgc	cctcaaccag	23760
gacggggcca	gcaacggctt	cgccgcggcg	aacgggaccg	cccagcgcaa	ggtgatccgg	23820
caggcgctac	gaaactgcgg	cctgaccccc	gccgacgtgg	acgccgtgga	ggcgacggc	23880
accggcacca	cgctcggcga	cccgatcgag	gccaacgccc	tgctggacac	ctacggccgt	23940
gaccgggatc	cggaccaccc	gctgtggctg	gggtcgggtga	agtcgaacat	cggccacacg	24000
caggcgggcg	cgggcgctac	cgggctgctc	aagatgggtg	tggcactgcg	ccacgaggaa	24060
ctgcccgcga	ccctgcacgt	cgacgagccc	accccgacag	tggactggtc	ctcgggagcg	24120
gtacgcctgg	cgacccgggg	ccggccgtgg	cggcgggggt	accggccgag	gcgggcccgg	24180
gtgtcggcgt	tcggcatcag	cgggaccaac	gcccacgtga	tcgtcgagga	ggcacccgag	24240
cggaccaccg	agcgcacctg	cggcggcgac	gtcggccccg	tcccgtcgt	ggtgtccgcc	24300
cggtcggcgg	cggcgctacg	ggcccaggcg	gcccaggctc	ccgagctggt	ggagggctcc	24360
gacgtcgggc	tggcgagggt	cgggcggagc	ctggccgtga	cccgggcgcg	acacgagcac	24420
cgggcgggcg	tgggtggcgt	gacccggggc	gagggcggtg	gggggctgcg	cgaggctcg	24480
gcggtcgaac	cgcgcggcga	ggacaccgtc	accggggctc	ccgagacgtc	cgggcgaccc	24540
gtcgtcttcc	tcttcccggg	acaggggtcc	cagtgggtcg	ggatgggcgc	ggagctgctg	24600
gactcggcac	cggcgttcgc	cgacacgata	cgcgcctgcg	acgaggcgat	ggcaccgttg	24660
caggactggt	cggtctccga	cgtgctccgg	caggagccgg	gggcaccggg	actggaccgg	24720
gtcgacgtgg	tgcagccggt	gctgttcgcg	gtgatgggtg	cgttggcgcg	gttgtggcag	24780
tcgtacgggg	tcacccccgc	tgcgggtggg	gggcactcgc	agggggagat	cgccgccgcc	24840
cacgtggcgg	gtgcgctctc	cctcgccgac	gcggcgaggc	tgggtggggg	ccgcagccgg	24900
ttgctgcggg	cgctgtccgg	gggcggcgcc	atgagcgccg	tcgcgctcgg	tgaggccgag	24960
gtacgcggcc	gactgcggtc	gtgggaggac	cggatctccg	tggccgccgt	caacggaccc	25020
cggtcgggtg	tgggtggcgg	ggaaccggag	gcgctgcggg	agtggggacg	ggagcgggag	25080
gccgagggcg	tacgggtccg	cgagatcgac	gtcgactacg	cctcgcactc	gccgcagatc	25140
gacagggctc	gtgacgaact	cctgacggtc	acgggggaga	tcgagccccg	gtcggcgagg	25200
atcaccttct	actcgacggt	cgacgtccgt	gctgtcgacg	gcaccgacct	ggacgcgggg	25260
tactgggtacc	gcaacctgcg	ggagacggtc	cggttcgccg	acgcgatgac	ccggttggcc	25320
gactcgggat	acgacgcggt	cgtcgaggtc	agcccgcata	cggtgggtgg	gtcggcggtc	25380
gccgagggcg	tcgaggaggc	aggtgtcgag	gacgccgtcg	tcgtcggcac	cctgtcccgg	25440
ggcgacggcg	gaccgggggc	gttcctgcgg	tcggcgggca	ccgcccactg	cgccgggtgt	25500
gacgtcgact	ggacgcccgc	cctcccggga	gctgcgacga	tcccgttgcc	gacgtacccg	25560
ttccaacgga	agccgtactg	gctgcggctc	tctgctcccg	ccccgcctc	ccacgatctc	25620
gcctaccggg	tgtcctggac	gccgatcacc	ccgcccgggg	acggcggtact	cgacggcgac	25680
tggctgggtg	tgcaccccgg	gggcagcacc	ggatgggtcg	acgggttggc	ggcgggcgtc	25740



accgccggcg	gtggccgggt	cgtcgcccac	ccggtggact	ccgtgaccte	ccggaccggc	25800
ctggccgagg	cgctcgcccc	gcgggacggc	acgttcgggg	gggtgctgtc	gtgggtggcg	25860
accgacgaac	ggcacgtcga	ggccgggtgcg	gtcgccctgc	tgaccctggc	gcaggcggtg	25920
ggtgacgccg	gaatcgacgc	accactgtgg	tgcctgacct	aggaggcggt	ccgtaccccc	25980
gtcgacgggtg	acctggcccc	accggcgacg	gcgcacctgc	acggtttcgc	ccaggtcgcc	26040
cggctggagc	tgccccggcg	cttcgggtggg	gtgctcgacc	tgccccccac	cgtcgacgcc	26100
gccgggacgc	gtctgggtcgc	ggcggtcctc	gccggcgggc	gcgaggacgt	cgtcgcccgc	26160
cgtggcgacc	gtctctacgg	ccgtcgccctg	gtcaggggcga	ccctgcgcgc	gccccggcggg	26220
gggttcaccc	cgcacggcac	cgctcctggtc	accggcgcg	ccggtccggt	gggagggtcg	26280
ctggccccgt	ggctcgccga	acgggggtgcc	acccgactcg	tcttgcccgg	cgcacacccg	26340
ggcgaggagt	tgttgaccgc	gateccggg	gccgggtgcca	ccgcccgtgg	gtgcgaaccg	26400
gaggcgagg	cactgcgtac	ggcgatcggc	ggggagtgtc	cgaccgcgct	cgtacacgcc	26460
gagacgttga	cgaacttcgc	cggcgtcgcc	gacgcgcacc	ccgaggactt	cgcgcgccacc	26520
gtcgcgggcga	agaccgcgct	gccgacggtc	ctggcgagg	tgctcgggcga	ccaccgcctc	26580
gaacgggagg	tctactgctc	gtcggtggcc	ggggtctggg	gtgggggtcg	catggccgcg	26640
tacgcgcgcg	gcagcgcccta	cctcgacgcc	ctggtcgagc	accgtcgcg	ccggggggcac	26700
gccagcgccct	cgggtggcctg	gaccccggtg	gccctgcccc	gcgcgggtcga	cgacgggtcg	26760
ctgcgcgagc	gcggcctgcg	cagcctcgac	gtggcccgac	ccctcgggac	gtgggaacgt	26820
ctgctccgcg	ccggtgcggt	gtcggtggcc	gtcgcccgac	tcgactggtc	ggtcttcaca	26880
gagggtttcg	cggccatccg	gccgaccccc	ctcttcgacg	aactcctcga	ccggcgcggg	26940
gaccccgacg	gcgcgcccgt	cgaccggccg	ggggagccgg	cgggcgagtg	gggtcgacga	27000
atcgcggcgc	tgtccccgca	ggaacagcgg	gagacgttgc	tgaccctcgt	cggcgagacg	27060
gtcgcgagg	tgctgggaca	cgagaccggc	accgagatca	acacccgtcg	ggccttcagc	27120
gaactcggcc	tcgactcgct	gggctcgatg	gccctgcgtc	agcgccctgg	ggcccgtacc	27180
ggcctgcgga	tgccggcctc	gctggtcttc	gaccacccga	cggtcaccgc	gctcgcgcg	27240
tacctgcgtc	gactggtcgt	cggggactcc	gacccgaccc	cggtaagggt	gttcggcccc	27300
accgacgagg	ccgaaccgct	cgcctgtggtc	ggcatcggtc	gccggttccc	cggcggcac	27360
gccacccccg	aggacctctg	gcgggtggtg	tccgagggca	cctccatcac	caccggattc	27420
cccaccgacc	ggggctggga	cctccggcgg	ctctaccacc	ccgaccggga	ccacccccgg	27480
accagctacg	tcgacagggg	gggattcctc	gacggggccc	cggacttcga	ccccgggttc	27540
ttcgggatca	ccccccgcga	ggcgctggcg	atggaccgcg	agcagcggtc	caccctggag	27600
atcgcggtgg	aggcggtgga	acgggcgggc	atcgaccggg	agaccctcct	cggcagcgac	27660
accggcgctc	tcgtcggcct	gaacggccag	tcctacctgc	aactgctgac	cggggagggt	27720
gaccggctca	acggctacca	ggggttgggc	aactcggcga	gcgtgctctc	cggccgtgtc	27780
gcctacacct	tcgggtggga	ggggcgcgcg	ctgacgggtg	acaccgcctg	ctcgctcctc	27840
ctggtcgcca	tcacacctgc	catgcagtcg	ctgcgtcggg	gtgagtgtc	gctggcggtg	27900
gccggcgggg	tgacgggtcat	ggccgacccg	tacaccttcg	tggacttcag	cgcacagcgg	27960
gggctcgccg	ccgacggggc	gtgcaaggcg	ttctccgcgc	aggccgacgg	gttcgcccctc	28020
gccgagggcg	tcgcggcgct	cgctcctcgaa	ccgttgctcca	aggcgcgggc	aaacggccac	28080
caggtgctgg	cgggtgctgcg	cggcagcgcc	gtcaaccagg	acggggccag	caacggcctc	28140
gccgccccga	acggggccgtc	gcaggaacgg	gtgatcaggc	aggccctgac	cgcctccggg	28200
ctgcgtcccc	ccgacgtcga	catgggtggag	gcgcacggga	cgggcaccca	actcggcgac	28260
ccgatcgagg	ccggggcgct	catcgcgcg	tacggccggg	accgggaccg	gccgctctgg	28320
ctgggctcgg	tgaagacgaa	catcgggccac	accagggccg	ccgcgggtgc	cgcgggggtg	28380
atcaaggcgg	tcctggcgat	gcggcacggc	gtactcccga	ggtcgctgca	cgcgcagcag	28440
ttgtccccgc	acatcgactg	ggcggaagg	aaggtcgagg	tgctccgcga	ggcacgacag	28500
tgcccccccc	gtgagcgccc	ccgcgcgcgc	ggggtgtcct	ccttcggcgt	cagcggggacc	28560
aacgcccacg	tcctcgtcga	ggaggcaccc	gccgaaccgg	accccgaaac	ggttcccgc	28620
gccccggggc	ggccccctgc	cttcgtcctg	cacggacgca	gcgtccagac	ggtccgggtc	28680
caggcgcgga	ccctcgccga	acacctgcgc	accacgggcc	accgggacct	cgcgcgacac	28740
gcccgtaccc	tggccaccgg	tcgcgcccgt	ttcgacgtcc	gggcccagct	gctcggcacc	28800
gaccgggagg	gtgtctgcgc	cgcctcgcac	gcgtggcg	aggatcgccc	ctcgcccgc	28860
gtcgtcgccc	cggcggtctt	cgcgcgccgt	acccccgtcc	tggtcttccc	cgggcagggg	28920
tcgcagtggg	tcggcatggc	ccgtgacctg	ctcgactcct	ccgagggtgt	cgcgcagtcg	28980
atggggccgg	gcgcgcaggc	gctgtcgccg	tacaccgact	gggacctgct	cgcagtggtc	29040
cgtggggctg	gcgacccccga	cccgtacgac	cgggtggacg	tgctccagcc	ggtgctgttc	29100
gcgggtgatg	tgtcgctggc	gcgggtgtgg	cagtcgtacg	gggtgactcc	gggtgcgggtg	29160
gtgggtcact	cgcagggggga	gatecgccgc	gcgcacgtgg	ctgggtgcgt	gtcggtggcc	29220
gacgcgcgca	gggtgggtggc	gttgcgacgc	cgggtgctgc	gggagctcga	cgcacagggc	29280
ggcatggtgt	cggtcggcac	ctcccgcgcc	gagttggact	cggtcctgcg	ccgggtgggac	29340
gggagggtcg	cgggtggcggc	ggtgaacgga	cccggcacgc	tcgtgggtggc	cggacccacc	29400

gccgaactgg	acgagttcct	cgcggtggcc	gaggcccgcg	agatgaggcc	gcgtcggatc	29460
gcggtgcgct	acgcgtcgca	ctccccggag	gtggcccggg	tcgaacagcg	gctcgccgcc	29520
gaactcggca	ccgtcaccgc	cgtcggcggc	acggtcccgc	tetactccac	cgccaccggg	29580
gacctcctcg	acaccacagc	catggacgcc	gggtactggt	accgcaacct	gcgccaaccg	29640
gtgctgttcg	agcacgccgt	ccgcagcctc	ctggagcggg	gattcgagac	gttcacgcag	29700
gtcagccccg	accctgtgct	gctgatggcg	gtcgaggaga	ccgccgagga	cgccgagcgc	29760
ccggtcaccg	gcgtgccgac	gctgcgccgc	gaccacgacg	ggccgtcggg	gttcctccgc	29820
aacctcctgg	gggcgcacgt	gcacgggggt	gacgtcgacc	tgcgtcgggc	ggtcgcccac	29880
ggccgcctgg	tcgacctgcc	cacctacccc	ttcgacaggg	agcggtctct	gcccagccg	29940
caccgcaggg	ccgacacctc	gtcgctgggg	gtccgtgact	cgacccaccc	gctgctgcac	30000
gccgcagtcg	acgtacccgg	tcacggcgga	gcggtgttca	ccgggcggct	ctcccccgac	30060
gagcagcagt	ggctgaccca	gcacgtgggt	ggtgggcgga	acctgggtgcc	cggcagtgtc	30120
ctggtcgacc	tcgcgctcac	cgccggggcc	gacgtcggcg	tgccggtgct	ggaggaactc	30180
gtcctgcagc	agccgctggt	gttgaccgcc	gccggtgcgt	tgctgcgcct	gtcggtcggc	30240
gccgccgacg	aggacggggc	gcggccgggt	gagatccacg	ccgccgagga	cgtctccgac	30300
ccggccgagg	cccgggtggc	ggcgtagcgc	accgggaccc	tcgccgtcgg	cgtggccggc	30360
ggcgccgggg	acggcacaca	gtggcccccg	cccggcgcca	ccgccctgac	gttgaccgac	30420
cactacgaca	ccctcgccga	actgggctac	gagtaggggc	cggcgttcca	ggcgctgcgc	30480
gccgcgtggc	agcacggcga	cgtggtctac	gcggagggtg	ccctcgacgc	cgtcgaggag	30540
gggtacgcgt	tcgacccggg	gctgctcgac	gccgtcgccc	agaccttcgg	cctgaccagt	30600
cgcgcccccg	ggaagctccc	cttcgcctgg	cggggcgtca	ccctgcacgc	caccggggcc	30660
actgcggtac	gggtgggtgg	gacccccgcc	ggaccggacg	cggtggccct	gcgggtcacc	30720
gacccgaccg	gtcagctcgt	cgccacgggt	gacgcctcgt	tcgtcaggga	cgccggggcg	30780
gatcgggacc	agccgcgcgg	ccgcgacggc	gacctgcacc	gcctggagtg	ggtacggctg	30840
gccaccccgg	acccgacccc	ggcggcgggt	gtgcacgtgg	cgcccgacgg	gctcgacgac	30900
ctgctgcgcg	ccggtgggtc	ggcaccacag	gccgtcgctg	tcgcctaccg	tcccgacggc	30960
gacgacccga	cggcgcgagg	ccgtcacggg	gtgctctggg	cgccacagct	cgtgcgccgt	31020
tggctcgacg	acgaccgggt	gcccgcaccc	acctgggtgg	tggccacgtc	cgcaggggtc	31080
gaggtctccc	ccggggacga	cgtgccgcgc	cccggggccg	ccgcctgtgt	gggggtgctg	31140
cgctgcgcgc	aggcggagtc	cccggaccgc	ttcgtgctcg	tcgacggcga	cccggagacg	31200
cccccggcgg	tgccggacaa	tcgcgagctc	gcggtccgtg	acggtgcggt	gttcgtgcca	31260
cggtgacgc	cgtcgcgcgg	tcctgtgcgc	gccgtcgccg	accgggcgta	ccggtgggtg	31320
cccggcaacg	gcgggtccat	cgaggcagtg	gccttcgccc	ccgtccccga	cgccgaccgg	31380
cccctggcgc	cggaggaggt	acgcgtcgcc	gtccgcgcga	ccggcgtgaa	cttcctgtgac	31440
gtcctgctcg	cgtcggcat	gtacccggaa	ccggccgaga	tgggcaccga	ggcgtccggt	31500
gtggtcaccg	aggtcgggtc	gggtgtccgg	cggttcaccc	ccggccaggc	ggtgacgggc	31560
ctgttccagg	gggccttcgg	gccggtggcg	gtcgccgacc	accggctcct	caccccggtc	31620
cccgcagggg	ggcgggcggg	ggacgcgcga	gccgtaccca	tcgcgttcac	caccgcccac	31680
tacgcgctgc	acgacctggc	cgggttgcag	gccgggcagt	ccgtgctggt	ccacgcgcgc	31740
gccggcgggg	tggggatggc	tgccgtcgcg	ttggcccgct	gggcgcgggg	ggaggtgttc	31800
gccacggcca	gcccggccaa	acacccgacg	ctgcgggcgc	tcggcctcga	cgacgaccac	31860
atcgccctcg	cccgggagag	cggtttcggt	gagcggttcg	ccgcgcgtac	cgggggggcg	31920
ggcgtcgacg	tggtcctgaa	ctcgctcacc	ggcgacctgc	tcgacgagtc	cgcgcggctg	31980
ctcgccgacg	gcgggggtct	cgtcgagatg	ggcaagaccg	acctgcggcc	ggcggagcag	32040
ttccggggcc	ggtacgtccc	gttcgacctg	gccgaggccg	gtcccgatcg	gctcggcgag	32100
atcctggagg	aggtcgtcgg	tctgctggcc	gccggtgccc	tcgaccgggt	gccggtgtcg	32160
gtgtgggagt	tgtcggcggc	cccggccgcg	ctcaccacaa	tgagccgggg	ccgacacgtg	32220
ggcaagctcg	tcctcaccca	gcccgccccc	gtgcaccccc	acggaacggt	gctggtcacc	32280
ggcgggaccg	gcacctggg	gcggctgggt	gcccgcaccc	tggtgaccgg	gcacggcgta	32340
ccccacctcc	tggtgggcag	ccggcgcggt	ccggcgggcc	cgggcgcggc	cgagctgcgc	32400
gccgacgtcg	aaggcctcgg	cgcgaccatc	gagatcgctg	cctgcgacac	cgccgaccgg	32460
gaggcgctcg	cggcgctgct	cgactcgatc	cccgcggacc	gtccgctgac	cggggtgggt	32520
cacaccgccc	gggtcctggc	cgacgggctg	gtcacctcca	tcgacgggac	cgccaccgat	32580
caggtcctgc	gggccaaggt	cgacgcggcg	tggcacctgc	acgacctgac	ccgggacgcg	32640
gacctgagct	tcttcgtgct	gttctcgtcg	gcggcgctcg	tgctggccgg	tcgggggacg	32700
ggcgtgtacg	cggcggccaa	cggggtcctc	aacgccctgg	ccggggcaacg	gcgggcccct	32760
ggactgcccg	cgaaggcgct	cggtgggggc	ctgtgggcgc	aggccagcga	gatgaccagc	32820
ggcctcggtg	accggatcgc	ccgtaccggg	gtcgccgcgc	tgccgaccga	gcgggcgctg	32880
gccctgttcg	acgcggctct	gcgcagcggc	ggggaggtgc	tgttcccgcct	gtctgtcgac	32940
aggtcggcgc	tgccgcgggc	cgagtacgtc	cccagaggtg	tgccgcggcg	ggtcgggtcc	33000
acgccacggg	ccgccaacag	ggccgagacc	ccgggcccgg	gcctgctcga	ccgtctcgtc	33060

ggtgcacccg	agaccgatca	ggtggccgcg	ctggccgagc	tggtccgctc	gcacgcggcg	33120
gcggtcgccg	gctacgactc	ggccgaccag	ctgcccgaac	gcaaggcggt	caaggacctc	33180
gggttcgact	cgctggcggc	ggtggagctg	cgcaaccggc	tcggcgtcac	caccggcgta	33240
cggtgcccc	gcacgctggt	gttcgaccac	ccgacaccgc	tggcgggtgg	cgaacacctg	33300
cggtcggagt	tgttcgccga	ctccgcgcgc	gacgtcgggg	tcggtgcgcg	cctcgacgac	33360
ctggaacggg	cgctcgacgc	cctgcccgcg	gcgaggggac	acgccgacgt	cggggcccgc	33420
ctggaggcgc	tgctgcgcgc	gtggcagagc	cgacgacccc	cggagaccga	gccagtgcg	33480
atcagtgcg	acgccagtga	cgacgagctg	ttctcgatgc	tcgacaggcg	tctcggcggg	33540
ggaggggagc	tctaggtgac	aggctcgattc	cgccccgcgg	cagtggaccg	taccgccctg	33600
acaggtecac	cggttcgcgc	tcgcctccca	caccgcacgg	ccgggggtatc	cacggaaggg	33660
atccgatgag	cgagagcagc	ggcatgaccg	aggaccgcct	ccggcgctat	ctcaagcgca	33720
ccgtcgccga	actcgactcg	gtgacaggtc	ggctcgacga	ggctcgagtac	cgggcccgcg	33780
aaccgatcgc	cgctcgtcgc	atggcctgcc	ggttcccccg	gggtgtggac	tcgccggagg	33840
cgttctggga	gttcattccgc	gacgggtggtg	acgcgatcgc	cgaggcgccc	acggaccgtg	33900
gctggccgcg	ggcaaccgcga	ccccgcctcg	gtgggtctcct	cgcgaggccg	ggcgcggttcg	33960
acgccgcctt	cttcggcatc	tcaccccgcg	aggcgctcgc	gacggacccc	cagcagcgcc	34020
tgatgctgga	gatctcctgg	gaggcggttg	agcgtgcggg	tttcgacccg	tcgagcctgc	34080
gcggcagcgc	cggtggcgctc	ttcaccgggtg	tcggtgcggt	ggactacgga	cccaggcccg	34140
acgaggcacc	cgaggagggtg	ctcggctacg	tcggcatcgg	caccgcctcc	agcgtcgctc	34200
ccggacgggt	ggcgtaacac	ctgggggttg	agggtccagc	cgtaaccgte	gacaccgcct	34260
gctcctccgg	gctcaccgcg	gtgcacctgg	cgatggagtc	gctgcgcgcg	gacgagtgc	34320
ccctggtcct	cgccgggtggg	gtcaccgtga	tgagcagccc	gggtgcgttc	accgagttcc	34380
gcagccaggg	cggtgtggcc	gaggacggcc	gctgcaaacc	gttctccgcg	gccgccgacg	34440
gcttcggggt	cgccgagggg	gccgggggtcc	tggtgctcca	acggctgtcc	gtcgcccggg	34500
ccgagggccg	gccgggtgctg	gccgtactgc	gtggctcggc	gatcaaccag	gacgggtgcca	34560
gcaacggggt	caccgcgcgc	agcggccccg	cccagcggcg	ggtgatcagg	caggcggttg	34620
agcggggcgc	gctgcgtccc	gtcgacgtgg	actacgtgga	ggcccacggc	accggcaccc	34680
ggctgggcga	tccgatcgag	gcgacacgccc	tgctcgacac	gtacgggtgcc	gaccgggaac	34740
ccggccgccc	gctctgggtc	ggatcggtga	agtccaacat	cggtcacacc	caggcgggcg	34800
cggggggtgg	cggggtgatg	aagaccgtgc	tggcgctgcg	gcatcgggag	atcccggcga	34860
cgttgcactt	cgacgagccc	tcgccgcacg	tcgactggga	ccgggggtgcg	gtgtcggtgg	34920
tgtecgagac	ccggccctgg	ccgggtgggg	agcgcccgcg	ccgggcgggg	gtgtcctcgt	34980
tcggcatcag	cggcaccaac	gcgacagtc	tcgtcgagga	ggcgccgagc	ccgcaggcgg	35040
ccgacctcga	cccgaacccc	ggcccggcaa	ccggagcgac	ccccggaaacg	gatgccgccc	35100
ccaccgcgga	gccgggtgcg	gaggcggtcg	cactgggtgtt	ctccgcgcgc	gacgagcggg	35160
ccctgcgcgc	ccaggcgggc	cggtcgcgcg	accgtctcac	cgacgacccg	gccccctcgt	35220
tgcgcgacac	cgccctcacc	ctggtcaccc	gccgtgccac	ctgggagcat	cgggcggtcg	35280
tcgtcggcgg	gggcgaggag	gtcctcgccg	gcctccgggc	cgtecgccggg	ggacgtcccg	35340
tcgacggagc	cgtcagcggg	cgggcgcgcg	ccggccgcgcg	ggtggtgctg	gtcttccccg	35400
ggcagggcgc	acagtggcag	ggcatggccc	gggacctgct	gcggcagtcg	ccgaccttcg	35460
cggagtccat	cgacgcctgc	gagcggggcg	tcgccccgca	cgtggactgg	tcgctgcgcg	35520
agggtgctcga	cggcgagcag	tcgttgagac	ccgtcgacgt	ggtgcagccg	gtgctgttcg	35580
cggtgatggt	gtcgttggcg	cggttgtggc	agtcgtacgg	ggtgactccg	ggtgcgggtg	35640
tgggtcactc	gcagggggag	atcgccgccc	cgcacgtggc	tggtgcgttg	tcgttgggcg	35700
acgccgcccag	ggtggtggcg	ttgcgcagcc	gggtgctgcg	ccgtctcggt	ggtcacggcg	35760
ggatggcgtc	gttcgggctc	caccccgcac	aggccgcgga	gcggatcgcg	cgcttcgcgg	35820
gtgcgctgac	tgtegcctcg	gtcaacggtc	cccgttcggt	ggtgctggcc	ggggagaacg	35880
gcccgttgga	cgagctgatc	gccgagtgcg	aggccgaggg	cgtgaccgcc	cgtcggatcc	35940
ccgtcgacta	cgccctcacac	tccccgcagg	tggagtcgct	gcgtgaggag	ctgctcgccg	36000
cactggccgg	ggtccgtccg	gtgtcggccg	ggatccccct	gtactcgacc	ctgaccggtc	36060
aggatcatcga	aacggcgacg	atggacgccc	actactgggt	cgccaaacctc	cgggagcccg	36120
tgcgcttcca	ggacgccacc	aggcagctcg	ccgaggcggg	gttcgacgcc	ttcgtcgagg	36180
tcagcccgcga	cccgggtgttg	acagtcggtg	tcgaggccac	cctcgaggca	gtgctgcccc	36240
ccgacgcgga	tccgtgtgtc	acaggcacc	tgccgcgcga	acgcggcggt	ctcgcgcagt	36300
tccacaccgc	gctcgccgag	gcgtacaccc	gggggggtgga	ggtcgactgg	cgtaaccgag	36360
tgggtgaggg	acgcccggtc	gacctgccgg	tctacccggt	ccaacgacag	aacttctggc	36420
tcccgggtccc	cctgggcccg	gtccccgcga	ccggcgacga	gtggcggttac	cagctcgccct	36480
ggcaccgccgt	cgacctcggg	cggtcctccc	tggccggacg	ggtcctggtg	gtgaccggag	36540
cggcagtaac	cccggcctgg	acggacgtgg	tccgcgacgg	cctggaacag	cgcggggcga	36600
ccgtcgtgtt	gtgcaccgcg	cagtcgcgcg	cccggatcgg	cgccgcactc	gacgcgctcg	36660
acggcacccg	cctgtccact	gtggtctctc	tgctcgcgct	cgccgagggc	ggtgctgtcg	36720



acgacccccag	cctggacacc	ctcgcgttgg	tccaggcgct	cggcgcagcc	gggatcgacg	36780
tccccctgtg	gctggtgacc	agggacgccg	ccgccgtgac	cgteggagac	gacgtcgatc	36840
cgccccaggc	catggtcggg	gggctcggcc	gggtggtggg	cgtggagtc	cccgcccggt	36900
ggggtggcct	ggtggacctg	cgcgaggccg	acgccgactc	ggcccggteg	ctggccgcca	36960
tactggccga	cccgcgcggc	gaggagcagt	tcgcgatccg	gcccgcaggg	gtcaccgtcg	37020
cccgctctcgt	cccggcaccg	gcccgcgcgg	cggttaccgg	gtggacgccg	cgcgggaccg	37080
tcctggtcac	cggcggcacc	ggcggcatcg	gcgcgcacct	ggcccgcctg	ctcgcgcgtg	37140
cgggcgccga	gcacctgggt	ctgctcaaca	ggcggggagc	ggaggcgggc	ggtgccggcg	37200
acctgcgtga	cgaactgggt	gcgctcggca	cgggagtcac	catcacggcc	tgcgacgtcg	37260
ccgaccgcga	ccggttggcg	gccgtcctcg	acgccgcacg	ggcgcagggg	cgggtggtca	37320
cggcggtggt	ccacgcgcgc	gggatctccc	ggtccacagc	ggtacaggag	ctgaccgaga	37380
gcgagttcac	cgagatcacc	gacgcgaagg	tgcgggggtac	ggcgaacctg	gccgaactct	37440
gtcccagagct	ggacgccctc	gtgctgttct	cctcgaacgc	ggcgggtgtg	ggcagcccgg	37500
ggctggcctc	ctacgcggcg	ggcaacgcct	tcctcgacgc	cttcgcccggt	cgtggtcggc	37560
gcagtgggct	gccggtcacc	tcgatcgctt	ggggtctgtg	ggccggggcag	aacatggccg	37620
gtaccgaggg	cggcgactac	ctgcgcagcc	agggcctgcg	cgccatggac	ccgcagcggg	37680
cgatcgagga	gctgcggacc	accctggacg	ccggggaccc	gtgggtgtcg	gtgggtggacc	37740
tggaccggga	gcggttcgtc	gaactgttca	ccgccgcccg	ccgccggccc	ctcttcgacg	37800
aactcgggtg	ggtccgcgcc	ggggccgagg	agaccgggtca	ggaatcggat	ctcgcgccgc	37860
ggctggcgct	gatgccggag	gccgaacgtc	acgagcatgt	cgcccggctg	gtccgagccg	37920
aggtggcagc	ggtgctgggc	cacggcacgc	cgacggtgat	cgagcgtgac	gtcgccttcc	37980
gtgacctggg	attcgactcc	atgaccgcgc	tcgacctgcg	gaaccggctc	gcggcgggtga	38040
ccgggggtccg	ggtggccacg	accatcgtct	tcgaccaccc	gacagtggac	cgcctcaccg	38100
cgcactacct	ggaacgactc	gtcgggtgagc	cggaggcgac	gaccccggtc	gcggcgggtcg	38160
tcccgcaggg	accgggggag	gccgacgagc	cgatcgcgat	cgtegggatg	gcctgccgcc	38220
tcgccgggtg	agtgcgtacc	cccgaccagt	tgtgggactt	catcgtcgcc	gacggcgacg	38280
cggtcaccga	gatgccgtcg	gaccgggtcct	gggacctcga	cgcgctgttc	gacccggacc	38340
ccgagcggga	cggcaccagc	tactcccggc	acggcgcggt	cctggacggg	gcggccgact	38400
tcgacgcggc	gttcttcggg	atctcgcgcg	gtgaggcggt	ggcgatggat	ccgcagcagc	38460
ggcaggtcct	ggagacgacg	tgggagctgt	tcgagaacgc	cggcatcgac	ccgcactccc	38520
tgcgcgggtac	ggacaccggg	gtcttcctcg	gcgctgcgta	ccaggggtac	ggccagaacg	38580
cgcaggtgcc	gaaggagagt	gaggggttacc	tgtcaccggg	tgggttcctcg	gcggtcgcct	38640
ccggtcggat	cgcgtacgtg	ttgggggttg	aggggcccgg	gatcactgtg	gacacggcgt	38700
gttcgctcgt	gcttgtggcg	ttgcacgtgg	cggccgggtc	gctgcgatcg	ggtgactgtg	38760
ggctcgcggg	ggcgggtggg	gtgtcgggtga	tggccgggtc	ggagggtgtc	accgagttct	38820
ccaggcaggg	cgcgctggcc	cccgcacggc	ggtgcaagcc	cttctccgac	caggccgacg	38880
ggttcggatt	cgcgcagggc	gtcgcgtgtg	tgtcctcgca	gcgggtgtcg	gtggcggtgc	38940
gggagggggc	tcgggtgttg	ggtgtggtgg	tgggttcggc	ggtgaatcag	gatggggcga	39000
gtaatgggtt	ggcggcgccg	tcgggggtgg	cgcagcagcg	ggtgattcgg	cgggcgtggg	39060
gtcgtgcggg	tgtgtcgggt	ggggatgtgg	gtgtggtgga	ggcgcagtgg	acggggacgc	39120
ggttggggga	tccggtggag	ttgggggctg	tgttggggac	gtatgggggt	ggtcgggggt	39180
gggtgggtcc	ggtggtgggt	ggttcgggtga	aggcgaatgt	gggtcatgtg	caggcggcgg	39240
cgggtgtggg	gggtgtgac	aagggtggtg	tgggggttgg	tcgggggttg	gtgggtccga	39300
tgggtgtgtc	gggtgggttg	tcgggggttg	tggattggtc	gtcgggtggg	ttgggtggtg	39360
cggatggggg	gcgggggtgg	ccgggtgggt	tggatggggg	gcgtcggggg	ggggtgtcgg	39420
cgtttggggg	gtcggggacg	aatgctcatg	tgggtggtgg	ggaggcgccg	gggtcgggtg	39480
tggggggcga	acggccgggt	gaggggtcgt	cgcgggggtt	ggtgggggtg	gctggtggtg	39540
tgggtgccgg	ggtgctgtcg	gcaaagaccg	aaaccgccct	gaccgagctc	gcccgcagac	39600
tgcacgacgc	cgtcgacgac	accgtcgccc	tcccggcggt	ggccgccacc	ctcgcaccg	39660
gacgcgcccc	cctgccctac	cgggcccggc	tgtggcccg	cgaccacgac	gaactgcgcg	39720
acaggctgcg	ggcggttcacc	actggttcgg	cggctcccgg	tgtggtgtcg	ggggtggcgt	39780
cgggtggtgg	tgtggtgttt	gtttttcctg	gtcagggtgg	tcagtgggtg	gggatggcgc	39840
gggggttgtt	gtcgggttcg	gtgtttgtgg	agtcgggtgg	ggagtgtgat	gcgggtggtg	39900
cgtcgggtgt	gggggttttc	gtgttggggg	tgttggaggg	tcggtcgggt	gcgcgctcgt	39960
tggatcgggt	ggatgtggtg	cagccgggtg	tgttcgtggt	gatggtgtcg	ttggcgcggt	40020
tgtggcggtg	gtgtgggggt	gtgcctgcgg	cgggtggtgg	tcattcgcag	ggggagatcg	40080
cggcggcggg	ggtggcgggg	gtgttgtcgg	tgggtgatgg	tgcgcgggtg	gtggcggtgc	40140
gggcgcgggc	gttgcggggc	ttggccggcc	acggcggcac	ggtctccctc	gcgggtctccg	40200
ccgaacgcgc	ccgggagctg	atcgaccctt	ggtccgaccg	gatctcgggt	gcggcggtca	40260
actccccgac	ctcgggtggtg	gtctcgggtg	acccacaggc	cctcgcgcgc	ctcgtcgcgc	40320
actgcgcgca	gaccggtgag	cgggccaaga	cgtgcctgtg	ggactacgcc	tcccactccg	40380

cccacgtcga	acagatccgc	gacacgatcc	tcaccgacct	ggccgacgtc	acggcgcgcc	40440
gacccgacgt	cgccctctac	tccacgtcgc	acggcgcccc	ggcgcgccgc	acggacatgg	40500
acgcccggta	ctggtacgac	aacctgcgct	caccgggtgcg	cttcgacgag	gccgtcgagg	40560
ccgcccgtcg	cgacggctac	cgggtcttcg	tcgagatgag	cccacacccg	gtcctcaccg	40620
ccgcggtgca	ggagatcgac	gacgagacgg	tggccatcgg	ctcgctgcac	cgggacacccg	40680
gcgagcgga	cctggtcgcc	gaactcgccc	gggcccacgt	gcacggcgta	ccagtggact	40740
ggcgggcgat	cctccccgcc	acccacccgg	ttccctgcc	gaactacccg	ttcgaggcga	40800
cccgtactg	gctcgccccg	acggcgggcg	accaggtcgc	cgaccacccg	taccgcgtcg	40860
actggcggcc	cctggccacc	accccgggcg	agctgtccgg	cagctacctc	gtcttcggcg	40920
acgccccgga	gaccctcggc	cacagcgtcg	agaaggccgg	cgggctcctc	gtcccgggtg	40980
ccgctcccga	ccgggagtc	ctcgcggtcg	ccctggacga	ggcgggccgga	cgactcgccg	41040
gtgtgtcttc	cttcgccccg	gacaccgcca	cccacctggc	ccggcaccga	ctcctcggcg	41100
aggccgacgt	cgaggcccca	ctctggctgg	tcaccagcgg	cggcgctcgca	ctcgacgacc	41160
acgacccgat	cgactgcgac	caggcaatgg	tgtgggggat	cggacgggtg	atgggtctgg	41220
agaccccga	ccggtggggc	ggcctgggtg	acgtgaccgt	cgaacccacc	gccgaggacg	41280
gggtgggtctt	cgccgcccctc	ctggccgccc	acgaccacga	ggaccaggtg	gcgctgcgcg	41340
acggcatccg	ccacggccga	cggctcgctc	gcgccccgct	gaccacccga	aacgccaggt	41400
ggacacccgc	gggcacggcg	ctcgctcacgg	gcgggtacggg	tgcctcgggc	ggccacgtcg	41460
cgcgggtacct	ggcccggtcc	gggggtgaccg	atctcgctct	gctcagcagg	agcgggccccg	41520
acgcacccgg	tgcgcgccga	ctggccgccc	aactggccga	cctcgggggc	gagccgagag	41580
tcgaggcggtg	cgacgtcacc	gacggggccac	gcctgcgcgc	cctggtgcag	gagctacggg	41640
aacaggaccg	gccggtccgg	atcgctcgctc	acaccgcagg	ggtgcccgac	tcccgtcccc	41700
tcgacccgat	cgacgaactg	gagtcgggtca	gcgcccgcga	ggtgaccggg	gcgcggtctgc	41760
tcgacgagct	ctgcccggac	gccgacacct	tcgtcctggt	ctcctcgggg	gcgggagtgt	41820
ggggtagcgc	gaacctgggc	gcgtacgcgg	cagccaacgc	ctacctggac	gccctggccc	41880
accgccgccc	ccaggcgggc	cgggcccgcga	cctcggtcgc	ctggggggcg	tgggcccggcg	41940
acggcatggc	caccggcgac	ctcgacgggc	tgacccgggc	cggctctgcg	gcgatggcac	42000
cggaccgggc	gctgcgcgcc	tgcaccaggc	ggtggaccac	ccacgacacc	tgtgtgtcgg	42060
tagccgacgt	cgactgggac	cgcttcgcgc	tgggtttcac	cgccgcccgg	cccagacccc	42120
tgatcgacga	actcgtcacc	tccgcgcgcg	tggccgcccc	caccgctgcg	gcggcccccg	42180
tcccggcgat	gaccgccgac	cagctactcc	agttcacgcg	ctcgcacgtg	gccgcgatcc	42240
tcggtcacca	ggacccgga	gcggctcggt	tggaccagcc	cttcaccgag	ctgggcttcg	42300
actcgctcac	cgccgtcggc	ctgcgcaacc	agctccagca	ggccaccggg	cggacgctgc	42360
ccgcccgcct	ggtgttccag	caccccacgg	tacgcagact	cgccgaccac	ctcgcgcagc	42420
agctcgacgt	cggcaccgcc	ccggtcgagg	cgacgggcag	cgtcctgcgg	gacggctacc	42480
ggcgggccgg	gcagaccggc	gacgtccggt	cgtacctgga	cctgctggcg	aacctgtcgg	42540
agtcccgga	gcggttcacc	gacgcggcga	gcctggggcg	acagctggaa	ctcgtcgacc	42600
tggccgacgg	atccggcccc	gtcactgtga	tctgttgccg	gggcaactgcg	gcgctctccg	42660
ggccgcacga	gttcgcccga	ctcgccctcg	cgtgcgcgcg	caccgtgccg	gtgcgcgccc	42720
tcgcgcaacc	cgggtacgag	gcgggtgaac	cggtgccggc	gtcgatggag	gcagtgtctg	42780
gggtgcaggc	ggacgcggtc	ctcgcggcac	agggcgacac	gccgttcgtg	ctgggtcggac	42840
actcggcggg	ggccctgatg	gcgtacgccc	tggcgaccga	gctggccgac	cggggccacc	42900
cgccacgtgg	cgtcgtgctc	ctcgacgtgt	acccacccgg	tcaccaggag	gcggtgcacg	42960
cctggctcgg	cgagctgacc	gccgccctgt	tcgaccacga	gaccgtacgg	atggacgaca	43020
cccggctcac	ggccctgggg	gcgtacgaca	ggctgaccgg	caggtggcgt	ccgagggaca	43080
ccggtctgcc	cacgctggtg	gtggccgcca	gcgagccgat	gggggagtgg	ccggacgacg	43140
gttggcagtc	cacgtggccg	ttcgggcacg	acagggtcac	ggtgcccggg	gaccacttet	43200
cgatgggtgca	ggagcacgcc	gacgcgatcg	cgcggcacat	cgacgcctgg	ttgagcgggg	43260
agagggcag	aacacgaccg	atcgcgccgt	gctgggcccga	cgactccaga	tgatccgggg	43320
actgtactgg	ggttacggca	gcaacggaga	cccgtacccg	atgctgttgt	gcgggcacga	43380
cgacgacccg	caccgctggt	accggggggt	gggcggatcc	gggggtccgg	gcagccgtac	43440
cgagacgtgg	gtgggtgaccg	accacgccac	cgcctgtcgg	gtgctcgacg	acccgacctt	43500
cacccggggc	accggccgga	cgcgggagtg	gatgcgggce	gcggggcgccc	cggcctcgac	43560
ctggggcgag	ccgttcctgt	acgtgcacgc	cgcgtcctgg	gacgccgaac	tgcgccgacc	43620
gcaggagggtg	gaggaccggc	tgacgggtct	cctgcctgcc	ccgggggacc	gcctggacct	43680
ggtccgcgac	ctcgccctgg	cgatggcgte	gcgggggggtc	ggcgcggacg	accccgacgt	43740
gctgcgcgcc	gcgtgggacg	cccgggtcgg	cctcgacgcc	cagctcacc	cgcagcccct	43800
ggcggtgacc	gaggcggcga	tcgcccgggt	gcccggggac	ccgcacccgg	gggcgctgtt	43860
caccgccgtc	gagatgacag	ccaccgcgtt	cgtcgacgcg	gtgctggcgg	tgaccgccac	43920
ggcgggggcg	gcccagcgtc	tcgcccacga	ccccgacgtc	gccgcccgtc	tcgtcgcgga	43980
ggtgctgcgc	ctgcatccga	cggcgcacct	ggaacggcgt	accgccggca	ccgagacggt	44040

ggtgggag	cacacggteg	cggcgggcga	cgaggtegtc	gtgggtggteg	ccgcccga	44100
ccgtgacgcg	gggggtctteg	ccgacccgga	ccgcctcgac	ccggaccggg	ccgacgcccga	44160
ccggggccctg	tccgcccagc	gcgggtcaccc	cggccgggtg	gaggagctgg	tgggtggctct	44220
gaccaccgcc	gcactgcgca	gcgtcgccaa	ggcgctgccc	ggtctcaccg	ccggtggccc	44280
ggtegtcagg	cgacgtcggt	caccggctct	gcgagccacc	gcccactgcc	cggtcgaaact	44340
ctgaggtgcc	tgcgatgcgc	gtcgtcttct	cctccatggc	cagcaagagc	cacctgttcg	44400
gtctcggttc	cctcgccctg	gccttcgcgc	cggcgggcca	cgaggtagcg	gtcgtcgct	44460
caccggctct	caccgacgac	atcacggcgg	ccggactgac	ggccgtaccg	gtcggcaccg	44520
acgtcgacct	tgtcgacttc	atgacccacg	ccgggtacga	catcatcgac	tacgtccgca	44580
gcctggactt	cagcgagcgg	gacccggcca	cctccacctg	ggaccacctg	ctcggcatgc	44640
agaccgtcct	caccccagacc	ttctacgccc	tgatgagccc	ggactcgctg	gtcgagggca	44700
tgatctcctt	ctgtcggtcg	tggcgacccg	actggtegtc	tggaccgcag	accttcgccc	44760
cgtegatcgc	ggcgacgggtg	accggcgtgg	cccacgcccg	actcctgtgg	ggacccgaca	44820
tcacggtagc	ggcccggcag	aagttcctcg	ggctgctgcc	cggacagccc	gccgcccacc	44880
gggaggaccc	cctcgccgag	tggctcacct	ggtctgtgga	gaggttcggc	ggccgggtgc	44940
cgcaggacgt	cgaggagctg	gtggtcgggc	agtggacgat	cgaccccgcc	ccggtcggga	45000
tgcgcctcga	caccgggctg	aggacgggtg	gcattgcgta	cgtecgactac	aacggcccgt	45060
cggtgggtgcc	ggactggctg	cacgacgagc	cgacccgccc	acgggtctgc	ctcaccctgg	45120
gcattctccag	ccgggagaaac	agcatcgggc	aggtctccgt	cgacgacctg	ttgggtgcgc	45180
tcgggtgacgt	cgacgcccag	atcatcgcca	cagtggacga	gcagcagctc	gaaggcgtcg	45240
cccacgtccc	ggccaacatc	cgtacggctg	ggttcgtccc	gatgcacgca	ctgctgccga	45300
cctgcgcggc	gacgggtgcac	cacggcggtc	ccggcagctg	gcacaccgcc	gccatccacg	45360
gcgtgccgca	ggtgatcctg	cccgacggct	gggacaccgg	ggtccgcgcc	cagcggaccg	45420
aggaccaggg	ggcgggcatc	gccctgccgg	tggccgagct	gacctccgac	cagctccgcg	45480
aggcgggtgcg	gcggggtcctg	gacgatcccg	ccttcaccgc	cggtgcggcg	cggatgcggg	45540
ccgacatgct	cgccgagccg	tcccccgccc	aggtegtcga	cgtctgtgcg	gggtggtcg	45600
gggaacggac	cgccgtcgga	tgagcaccga	cgccaccac	gtccggctcg	gccgggtgcgc	45660
cctgctgacc	agccggctct	ggctgggtac	ggcagccctc	gccggccagg	acgacgccga	45720
cgcagtagcgc	ctgctcgacc	acgcccgttc	ccggggcgctc	aactgcctcg	acaccgccga	45780
cgacgactct	gcgtcgacca	gtgcccaggt	cgccgaggag	tcggtcggcc	ggtggttggc	45840
cggggacacc	ggtcggcggg	aggagaccgt	cctgtcggtg	acggtggtg	tcccaccggg	45900
cgggcaggtc	ggcggggggc	gcctctccgc	ccggcagatc	atcgccctct	gtgagggctc	45960
cctgcggcgt	ctcggtgtcg	accacgtcga	cgtecttcac	ctgccccggg	tggaccgggt	46020
ggagccgtgg	gacgaggtct	ggcaggcggt	ggacgcctc	gtggccgcgc	gaaaggtctg	46080
ttacgtcggg	tcgtcgggct	tccccggatg	gcacatcgctc	gccgcccagg	agcacgccgt	46140
ccgcccgtcac	cgcctcggcc	tgggtgccca	ccagtgtcgg	tacgacctga	cgtegcgcca	46200
tcccgaactg	gaggtcctgc	ccgcccgcga	ggcgtacggg	ctcggggtct	tcgccaggcc	46260
gacccgcctc	ggcgggtctgc	tcggcgggca	cggteccggg	gccgcagccg	cacgggctgc	46320
gggacagccg	acggcactgc	gctcggcggt	ggaggcgtag	gaggtgttct	gcagagacct	46380
cggcgagcac	cccgcggagg	tcgcaactgg	gtgggtgctg	tcccggcccc	gtgtggcggg	46440
ggcggtegtc	ggtgcgcgga	cgcccggacg	gctcgactcc	gcgtccgcgc	cctgcggcgt	46500
cgcctcggc	gcgacggaac	tcaccgcct	ggacgggatc	ttccccgggg	tcgccgcagc	46560
aggggcccgg	ccggaggcgt	ggctacgggtg	agagcccgc	cctgacctgc	gggaaccctg	46620
gtcgggtgcg	cgggacggcc	gccgcggctc	ccgccccggg	cagccgggtg	gggtgagccg	46680
cagcaggctc	ggcgccaccg	actcggccac	ctccccgacg	tggtcggcga	ggtagaagtg	46740
cccgcgggg	aaggtccggg	tacggccggg	gactaccgag	tacggcagcc	agcgttgggc	46800
gtcctccacc	gtcgtcaacg	ggtcgggtgc	accgcagagg	gtgggtgatgc	cggcccgcag	46860
cggcgggccc	gcctgccagg	cgtaggagcg	cagcaccggg	tggtcggccc	gcagcaccgg	46920
cagcgacatg	tccaacagcc	cctggteggc	caatgcggcc	tcgctgacct	cgagcctgcg	46980
catctgctcg	acgagtcctg	cctcgtcggg	caggtcgggtg	cgccgctcgt	ggaccggggg	47040
ggcgggtctgc	ccggagacga	acaaccgcag	cggtegcacc	cccggacgag	cctccaggcg	47100
acgggcgggtc	tcgtaggcga	ccaggggcgcc	catgctgtga	ccgaacaggg	cgaacggaac	47160
ctcggcgacg	aggtcgcgca	gcacggccgc	gacctcgtcg	gcgatctccc	cggcgggtgcc	47220
gagagcccgc	tcgtcacgtc	ggtectgccg	gcccgggtac	tgacccgccc	acacgtcgac	47280
ctccggggcc	agtgcggggg	cgaggtecgag	gtacgagtcg	gcggcggtc	ccgcgtgcgc	47340
gaagcagtag	agccggggcc	ggtgtccgtc	ggcggacccg	aaccgcccga	accaggtgtt	47400
catcggtgtc	tcattccgttc	ggtegcacccg	gcaggtgggtc	gatgcgcgcg	agcaggagcg	47460
accgcccga	gacaacctcg	tcggagggga	agcccagcga	cagcttcggg	aagcggtcga	47520
acaggggccc	cagggcgacc	tctccctcca	gcttgccag	cgggcggccc	atgcagtagt	47580
ggatgccgtg	cccgaagggtg	aggtgtcccc	ggctgtccct	ggtgacgtcg	aaccggtcgg	47640
ggteggggaa	ctgtcccggg	tcgcgggttg	ccgccccgtt	ggcgatcagg	acgggtgctgt	47700



```

acgccgggat cgtcaccgcc cccatctcca cctcggcggt ggcgaaccgg gtggtggtct 47760
ccggtggggc ctggtagcgc aggatctcct ccaccgctcc gggcagcagt gccgggtcct 47820
tccggaccag cgcgagctgg tcgggggtggg tcagcagcag gtaggtgccg atcccgatga 47880
ggctcaccga cgcctcgaat cccgccagca gcagcaccag cgcgatggag gtgagttcgt 47940
cgcggtcgag ccggtcggcg tcgtcgtcct ggaccggat c 47981

```

&lt;210&gt; 2

&lt;211&gt; 48

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 2

```

Met Gly Asp Arg Val Asn Gly His Ala Thr Pro Glu Ser Thr Gln Ser
1           5           10           15
Ala Ile Arg Phe Leu Thr Arg His Gly Gly Pro Pro Thr Ala Thr Asp
20           25           30
Asp Val His Asp Trp Leu Ala His Arg Ala Ala Glu His Arg Leu Glu
35           40           45

```

&lt;210&gt; 3

&lt;211&gt; 377

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 3

```

Met Ala Val Gly Asp Arg Arg Arg Leu Gly Arg Glu Leu Gln Met Ala
1           5           10           15
Arg Gly Leu Tyr Trp Gly Phe Gly Ala Asn Gly Asp Leu Tyr Ser Met
20           25           30
Leu Leu Ser Gly Arg Asp Asp Asp Pro Trp Thr Trp Tyr Glu Arg Leu
35           40           45
Arg Ala Ala Gly Arg Gly Pro Tyr Ala Ser Arg Ala Gly Thr Trp Val
50           55           60
Val Gly Asp His Arg Thr Ala Ala Glu Val Leu Ala Asp Pro Gly Phe
65           70           75           80
Thr His Gly Pro Pro Asp Ala Ala Arg Trp Met Gln Val Ala His Cys
85           90           95
Pro Ala Ala Ser Trp Ala Gly Pro Phe Arg Glu Phe Tyr Ala Arg Thr
100          105          110
Glu Asp Ala Ala Ser Val Thr Val Asp Ala Asp Trp Leu Gln Gln Arg
115          120          125
Cys Ala Arg Leu Val Thr Glu Leu Gly Ser Arg Phe Asp Leu Val Asn
130          135          140
Asp Phe Ala Arg Glu Val Pro Val Leu Ala Leu Gly Thr Ala Pro Ala
145          150          155          160
Leu Lys Gly Val Asp Pro Asp Arg Leu Arg Ser Trp Thr Ser Ala Thr
165          170          175
Arg Val Cys Leu Asp Ala Gln Val Ser Pro Gln Gln Leu Ala Val Thr
180          185          190
Glu Gln Ala Leu Thr Ala Leu Asp Glu Ile Asp Ala Val Thr Gly Gly
195          200          205
Arg Asp Ala Ala Val Leu Val Gly Val Val Ala Glu Leu Ala Ala Asn
210          215          220
Thr Val Gly Asn Ala Val Leu Ala Val Thr Glu Leu Pro Glu Leu Ala
225          230          235          240
Ala Arg Leu Ala Asp Asp Pro Glu Thr Ala Thr Arg Val Val Thr Glu
245          250          255
Val Ser Arg Thr Ser Pro Gly Val His Leu Glu Arg Arg Thr Ala Ala
260          265          270
Ser Asp Arg Arg Val Gly Gly Val Asp Val Pro Thr Gly Gly Glu Val
275          280          285

```

Thr Val Val Val Ala Ala Ala Asn Arg Asp Pro Glu Val Phe Thr Asp  
 290 295 300  
 Pro Asp Arg Phe Asp Val Asp Arg Gly Gly Asp Ala Glu Ile Leu Ser  
 305 310 315 320  
 Ser Arg Pro Gly Ser Pro Arg Thr Asp Leu Asp Ala Leu Val Ala Thr  
 325 330 335  
 Leu Ala Thr Ala Ala Leu Arg Ala Ala Ala Pro Val Leu Pro Arg Leu  
 340 345 350  
 Ser Arg Ser Gly Pro Val Ile Arg Arg Arg Ser Pro Val Ala Arg  
 355 360 365  
 Gly Leu Ser Arg Cys Pro Val Glu Leu  
 370 375

<210> 4  
 <211> 436  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 4  
 Met Arg Val Val Phe Ser Ser Met Ala Val Asn Ser His Leu Phe Gly  
 1 5 10 15  
 Leu Val Pro Leu Ala Ser Ala Phe Gln Ala Ala Gly His Glu Val Arg  
 20 25 30  
 Val Val Ala Ser Pro Ala Leu Thr Asp Asp Val Thr Gly Ala Gly Leu  
 35 40 45  
 Thr Ala Val Pro Val Gly Asp Asp Val Glu Leu Val Glu Trp His Ala  
 50 55 60  
 His Ala Gly Gln Asp Ile Val Glu Tyr Met Arg Thr Leu Asp Trp Val  
 65 70 75 80  
 Asp Gln Ser His Thr Thr Met Ser Trp Asp Asp Leu Leu Gly Met Gln  
 85 90 95  
 Thr Thr Phe Thr Pro Thr Phe Phe Ala Leu Met Ser Pro Asp Ser Leu  
 100 105 110  
 Ile Asp Gly Met Val Glu Phe Cys Arg Ser Trp Arg Pro Asp Trp Ile  
 115 120 125  
 Val Trp Glu Pro Leu Thr Phe Ala Ala Pro Ile Ala Ala Arg Val Thr  
 130 135 140  
 Gly Thr Pro His Ala Arg Met Leu Trp Gly Pro Asp Val Ala Thr Arg  
 145 150 155 160  
 Ala Arg Gln Ser Phe Leu Arg Leu Leu Ala His Gln Glu Val Glu His  
 165 170 175  
 Arg Glu Asp Pro Leu Ala Glu Trp Phe Asp Trp Thr Leu Arg Arg Phe  
 180 185 190  
 Gly Asp Asp Pro His Leu Ser Phe Asp Glu Glu Leu Val Leu Gly Gln  
 195 200 205  
 Trp Thr Val Asp Pro Ile Pro Glu Pro Leu Arg Ile Asp Thr Gly Val  
 210 215 220  
 Arg Thr Val Gly Met Arg Tyr Val Pro Tyr Asn Gly Pro Ser Val Val  
 225 230 235 240  
 Pro Ala Trp Leu Leu Arg Glu Pro Glu Arg Arg Val Cys Leu Thr  
 245 250 255  
 Leu Gly Gly Ser Ser Arg Glu His Gly Ile Gly Gln Val Ser Ile Gly  
 260 265 270  
 Glu Met Leu Asp Ala Ile Ala Asp Ile Asp Ala Glu Phe Val Ala Thr  
 275 280 285  
 Phe Asp Asp Gln Gln Leu Val Gly Val Gly Ser Val Pro Ala Asn Val  
 290 295 300  
 Arg Thr Ala Gly Phe Val Pro Met Asn Val Leu Leu Pro Thr Cys Ala  
 305 310 315 320  
 Ala Thr-Val His His Gly Gly Thr Gly Ser Trp Leu Thr Ala Ala Ile  
 325 330 335

His Gly Val Pro Gln Ile Ile Leu Ser Asp Ala Asp Thr Glu Val His  
 340 345 350  
 Ala Lys Gln Leu Gln Asp Leu Gly Ala Gly Leu Ser Leu Pro Val Ala  
 355 360 365  
 Gly Met Thr Ala Glu His Leu Arg Gly Ala Ile Glu Arg Val Leu Asp  
 370 375 380  
 Glu Pro Ala Tyr Arg Leu Gly Ala Glu Arg Met Arg Asp Gly Met Arg  
 385 390 395 400  
 Thr Asp Pro Ser Pro Ala Gln Val Val Gly Ile Cys Gln Asp Leu Ala  
 405 410 415  
 Ala Asp Arg Ala Ala Arg Gly Arg Gln Pro Arg Arg Thr Ala Glu Pro  
 420 425 430  
 His Leu Pro Arg  
 435

&lt;210&gt; 5

&lt;211&gt; 390

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 5

Met Val Thr Ser Thr Asn Leu Asp Thr Thr Ala Arg Pro Ala Leu Asn  
 1 5 10 15  
 Ser Leu Thr Gly Met Arg Phe Val Ala Ala Phe Leu Val Phe Phe Thr  
 20 25 30  
 His Val Leu Ser Arg Leu Ile Pro Asn Ser Tyr Val Tyr Ala Asp Gly  
 35 40 45  
 Leu Asp Ala Phe Trp Gln Thr Thr Gly Arg Val Gly Val Ser Phe Phe  
 50 55 60  
 Phe Ile Leu Ser Gly Phe Val Leu Thr Trp Ser Ala Arg Ala Ser Asp  
 65 70 75 80  
 Ser Val Trp Ser Phe Trp Arg Arg Arg Val Cys Lys Leu Phe Pro Asn  
 85 90 95  
 His Leu Val Thr Ala Phe Ala Ala Val Val Leu Phe Leu Val Thr Gly  
 100 105 110  
 Gln Ala Val Ser Gly Glu Ala Leu Ile Pro Asn Leu Leu Leu Ile His  
 115 120 125  
 Ala Trp Phe Pro Ala Leu Glu Ile Ser Phe Gly Ile Asn Pro Val Ser  
 130 135 140  
 Trp Ser Leu Ala Cys Glu Ala Phe Phe Tyr Leu Cys Phe Pro Leu Phe  
 145 150 155 160  
 Leu Phe Trp Ile Ser Gly Ile Arg Pro Glu Arg Leu Trp Ala Trp Ala  
 165 170 175  
 Ala Val Val Phe Ala Ala Ile Trp Ala Val Pro Val Val Ala Asp Leu  
 180 185 190  
 Leu Leu Pro Ser Ser Pro Pro Leu Ile Pro Gly Leu Glu Tyr Ser Ala  
 195 200 205  
 Ile Gln Asp Trp Phe Leu Tyr Thr Phe Pro Ala Thr Arg Ser Leu Glu  
 210 215 220  
 Phe Ile Leu Gly Ile Ile Leu Ala Arg Ile Leu Ile Thr Gly Arg Trp  
 225 230 235 240  
 Ile Asn Val Gly Leu Leu Pro Ala Val Leu Leu Phe Pro Val Phe Phe  
 245 250 255  
 Val Ala Ser Leu Phe Leu Pro Gly Val Tyr Ala Ile Ser Ser Ser Met  
 260 265 270  
 Met Ile Leu Pro Leu Val Leu Ile Ile Ala Ser Gly Ala Thr Ala Asp  
 275 280 285  
 Leu Gln Gln Lys Arg Thr Phe Met Arg Asn Arg Val Met Val Trp Leu  
 290 295 300  
 Gly Asp Val Ser Phe Ala Leu Tyr Met Val His Phe Leu Val Ile Val  
 305 310 315 320

Tyr Gly Ala Asp Leu Leu Gly Phe Ser Gln Thr Glu Asp Ala Pro Leu  
 325 330 335  
 Gly Leu Ala Leu Phe Met Ile Ile Pro Phe Leu Ala Val Ser Leu Val  
 340 345 350  
 Leu Ser Trp Leu Leu Tyr Arg Phe Val Glu Leu Pro Val Met Arg Asn  
 355 360 365  
 Trp Ala Arg Pro Ala Ser Ala Arg Arg Lys Pro Ala Thr Glu Pro Glu  
 370 375 380  
 Gln Thr Pro Ser Arg Arg  
 385 390

&lt;210&gt; 6

&lt;211&gt; 374

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 6

Met Thr Thr Tyr Val Trp Ser Tyr Leu Leu Glu Tyr Glu Arg Glu Arg  
 1 5 10 15  
 Ala Asp Ile Leu Asp Ala Val Gln Lys Val Phe Ala Ser Gly Ser Leu  
 20 25 30  
 Ile Leu Gly Gln Ser Val Glu Asn Phe Glu Thr Glu Tyr Ala Arg Tyr  
 35 40 45  
 His Gly Ile Ala His Cys Val Gly Val Asp Asn Gly Thr Asn Ala Val  
 50 55 60  
 Lys Leu Ala Leu Glu Ser Val Gly Val Gly Arg Asp Asp Glu Val Val  
 65 70 75 80  
 Thr Val Ser Asn Thr Ala Ala Pro Thr Val Leu Ala Ile Asp Glu Ile  
 85 90 95  
 Gly Ala Arg Pro Val Phe Val Asp Val Arg Asp Glu Asp Tyr Leu Met  
 100 105 110  
 Asp Thr Asp Leu Val Glu Ala Ala Val Thr Pro Arg Thr Lys Ala Ile  
 115 120 125  
 Val Pro Val His Leu Tyr Gly Gln Cys Val Asp Met Thr Ala Leu Arg  
 130 135 140  
 Glu Leu Ala Asp Arg Arg Gly Leu Lys Leu Val Glu Asp Cys Ala Gln  
 145 150 155 160  
 Ala His Gly Ala Arg Arg Asp Gly Arg Leu Ala Gly Thr Met Ser Asp  
 165 170 175  
 Ala Ala Ala Phe Ser Phe Tyr Pro Thr Lys Val Leu Gly Ala Tyr Gly  
 180 185 190  
 Asp Gly Gly Ala Val Val Thr Asn Asp Asp Glu Thr Ala Arg Ala Leu  
 195 200 205  
 Arg Arg Leu Arg Tyr Tyr Gly Met Glu Glu Val Tyr Tyr Val Thr Arg  
 210 215 220  
 Thr Pro Gly His Asn Ser Arg Leu Asp Glu Val Gln Ala Glu Ile Leu  
 225 230 235 240  
 Arg Arg Lys Leu Thr Arg Leu Asp Ala Tyr Val Ala Gly Arg Arg Ala  
 245 250 255  
 Val Ala Gln Arg Tyr Val Asp Gly Leu Ala Asp Leu Gln Asp Ser His  
 260 265 270  
 Gly Leu Glu Leu Pro Val Val Thr Asp Gly Asn Glu His Val Phe Tyr  
 275 280 285  
 Val Tyr Val Val Arg His Pro Arg Arg Asp Glu Ile Ile Lys Arg Leu  
 290 295 300  
 Arg Asp Gly Tyr Asp Ile Ser Leu Asn Ile Ser Tyr Pro Trp Pro Val  
 305 310 315 320  
 His Thr Met Thr Gly Phe Ala His Leu Gly Val Ala Ser Gly Ser Leu  
 325 330 335  
 Pro Val Thr Glu Arg Leu Ala Gly Glu Ile Phe Ser Leu Pro Met Tyr  
 340 345 350

Pro Ser Leu Pro His Asp Leu Gln Asp Arg Val Ile Glu Ala Val Arg  
           355                          360                          365  
 Glu Val Ile Thr Gly Leu  
           370

&lt;210&gt; 7

&lt;211&gt; 257

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 7

Met Pro Asn Ser His Ser Thr Thr Ser Ser Thr Asp Val Ala Pro Tyr  
   1                          5                          10                          15  
 Glu Arg Ala Asp Ile Tyr His Asp Phe Tyr His Gly Arg Gly Lys Gly  
                           20                          25                          30  
 Tyr Arg Ala Glu Ala Asp Ala Leu Val Glu Val Ala Arg Lys His Thr  
                           35                          40                          45  
 Pro Gln Ala Ala Thr Leu Leu Asp Val Ala Cys Gly Thr Gly Ser His  
           50                          55                          60  
 Leu Val Glu Leu Ala Asp Ser Phe Arg Glu Val Val Gly Val Asp Leu  
   65                          70                          75                          80  
 Ser Ala Ala Met Leu Ala Thr Ala Ala Arg Asn Asp Pro Gly Arg Glu  
                           85                          90                          95  
 Leu His Gln Gly Asp Met Arg Asp Phe Ser Leu Asp Arg Arg Phe Asp  
                           100                          105                          110  
 Val Val Thr Cys Met Phe Ser Ser Thr Gly Tyr Leu Val Asp Glu Ala  
                           115                          120                          125  
 Glu Leu Asp Arg Ala Val Ala Asn Leu Ala Gly His Leu Ala Pro Gly  
           130                          135                          140  
 Gly Thr Leu Val Val Glu Pro Trp Trp Phe Pro Glu Thr Phe Arg Pro  
   145                          150                          155                          160  
 Gly Trp Val Gly Ala Asp Leu Val Thr Ser Gly Asp Arg Arg Ile Ser  
                           165                          170                          175  
 Arg Met Ser His Thr Val Pro Ala Gly Leu Pro Asp Arg Thr Ala Ser  
                           180                          185                          190  
 Arg Met Thr Ile His Tyr Thr Val Gly Ser Pro Glu Ala Gly Ile Glu  
                           195                          200                          205  
 His Phe Thr Glu Val His Val Met Thr Leu Phe Ala Arg Ala Ala Tyr  
           210                          215                          220  
 Glu Gln Ala Phe Gln Arg Ala Gly Leu Ser Cys Ser Tyr Val Gly His  
   225                          230                          235                          240  
 Asp Leu Phe Ser Pro Gly Leu Phe Val Gly Val Ala Ala Glu Pro Gly  
                           245                          250                          255  
 Arg

&lt;210&gt; 8

&lt;211&gt; 201

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 8

Met Arg Val Glu Glu Leu Gly Ile Glu Gly Val Phe Thr Phe Thr Pro  
   1                          5                          10                          15  
 Gln Thr Phe Ala Asp Glu Arg Gly Val Phe Gly Thr Ala Tyr Gln Glu  
                           20                          25                          30  
 Asp Val Phe Val Ala Ala Leu Gly Arg Pro Leu Phe Pro Val Ala Gln  
                           35                          40                          45  
 Val Ser Thr Thr Arg Ser Arg Arg Gly Val Val Arg Gly Val His Phe  
           50                          55                          60  
 Thr Thr Met Pro Gly Ser Met Ala Lys Tyr Val Tyr Cys Ala Arg Gly

65                                      70                                      75                                      80  
 Arg Ala Met Asp Phe Ala Val Asp Ile Arg Pro Gly Ser Pro Thr Phe  
    85                                      90                                      95  
 Gly Arg Ala Glu Pro Val Glu Leu Ser Ala Glu Ser Met Val Gly Leu  
    100                                      105                                      110  
 Tyr Leu Pro Val Gly Met Gly His Leu Phe Val Ser Leu Glu Asp Asp  
    115                                      120                                      125  
 Thr Thr Leu Val Tyr Leu Met Ser Ala Gly Tyr Val Pro Asp Lys Glu  
    130                                      135                                      140  
 Arg Ala Val His Pro Leu Asp Pro Glu Leu Ala Leu Pro Ile Pro Ala  
 145                                      150                                      155                                      160  
 Asp Leu Asp Leu Val Met Ser Glu Arg Asp Arg Val Ala Pro Thr Leu  
    165                                      170                                      175  
 Arg Glu Ala Arg Asp Gln Gly Ile Leu Pro Asp Tyr Ala Ala Cys Arg  
    180                                      185                                      190  
 Ala Ala Ala His Arg Val Val Arg Thr  
    195                                      200

&lt;210&gt; 9

&lt;211&gt; 328

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 9

Met Val Val Leu Gly Ala Ser Gly Phe Leu Gly Ser Ala Val Thr His  
 1                                      5                                      10                                      15  
 Ala Leu Ala Asp Leu Pro Val Arg Val Arg Leu Val Ala Arg Arg Glu  
    20                                      25                                      30  
 Val Val Val Pro Ser Gly Ala Val Ala Asp Tyr Glu Thr His Arg Val  
    35                                      40                                      45  
 Asp Leu Thr Glu Pro Gly Ala Leu Ala Glu Val Val Ala Asp Ala Arg  
    50                                      55                                      60  
 Ala Val Phe Pro Phe Ala Ala Gln Ile Arg Gly Thr Ser Gly Trp Arg  
 65                                      70                                      75                                      80  
 Ile Ser Glu Asp Asp Val Val Ala Glu Arg Thr Asn Val Gly Leu Val  
    85                                      90                                      95  
 Arg Asp Leu Ile Ala Val Leu Ser Arg Ser Pro His Ala Pro Val Val  
    100                                      105                                      110  
 Val Phe Pro Gly Ser Asn Thr Gln Val Gly Arg Val Thr Ala Gly Arg  
    115                                      120                                      125  
 Val Ile Asp Gly Ser Glu Gln Asp His Pro Glu Gly Val Tyr Asp Arg  
    130                                      135                                      140  
 Gln Lys His Thr Gly Glu Gln Leu Leu Lys Glu Ala Thr Ala Ala Gly  
 145                                      150                                      155                                      160  
 Ala Ile Arg Ala Thr Ser Leu Arg Leu Pro Pro Val Phe Gly Val Pro  
    165                                      170                                      175  
 Ala Ala Gly Thr Ala Asp Asp Arg Gly Val Val Ser Thr Met Ile Arg  
    180                                      185                                      190  
 Arg Ala Leu Thr Gly Gln Pro Leu Thr Met Trp His Asp Gly Thr Val  
    195                                      200                                      205  
 Arg Arg Glu Leu Leu Tyr Val Thr Asp Ala Ala Arg Ala Phe Val Thr  
    210                                      215                                      220  
 Ala Leu Asp His Ala Asp Ala Leu Ala Gly Arg His Phe Leu Leu Gly  
 225                                      230                                      235                                      240  
 Thr Gly Arg Ser Trp Pro Leu Gly Glu Val Phe Gln Ala Val Ser Arg  
    245                                      250                                      255  
 Ser Val Ala Arg His Thr Gly Glu Asp Pro Val Pro Val Val Ser Val  
    260                                      265                                      270  
 Pro Pro Pro Ala His Met Asp Pro Ser Asp Leu Arg Ser Val Glu Val  
    275                                      280                                      285  
 Asp Pro Ala Arg Phe Thr Ala Val Thr Gly Trp Arg Ala Thr Val Thr



290 295 300  
 Met Ala Glu Ala Val Asp Arg Thr Val Ala Ala Leu Ala Pro Arg Arg  
 305 310 315 320  
 Ala Ala Ala Pro Ser Glu Pro Ser  
 325

<210> 10  
 <211> 330  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 10  
 Met Gly Thr Thr Gly Ala Gly Ser Ala Arg Val Arg Val Gly Arg Ser  
 1 5 10 15  
 Ala Leu His Thr Ser Arg Leu Trp Leu Gly Thr Val Asn Phe Ser Gly  
 20 25 30  
 Arg Val Thr Asp Asp Asp Ala Leu Arg Leu Met Asp His Ala Leu Glu  
 35 40 45  
 Arg Gly Val Asn Cys Ile Asp Thr Ala Asp Ile Tyr Gly Trp Arg Leu  
 50 55 60  
 Tyr Lys Gly His Thr Glu Glu Leu Val Gly Arg Trp Phe Ala Gln Gly  
 65 70 75 80  
 Gly Gly Arg Arg Glu Glu Thr Val Leu Ala Thr Lys Val Gly Ser Glu  
 85 90 95  
 Met Ser Glu Arg Val Asn Asp Gly Gly Leu Ser Ala Arg His Ile Val  
 100 105 110  
 Ala Ala Cys Glu Asn Ser Leu Arg Arg Leu Gly Val Asp His Ile Asp  
 115 120 125  
 Ile Tyr Gln Thr His His Ile Asp Arg Ala Ala Pro Trp Asp Glu Val  
 130 135 140  
 Trp Gln Ala Ala Glu His Leu Val Gly Ser Gly Lys Val Gly Tyr Val  
 145 150 155 160  
 Gly Ser Ser Asn Leu Ala Gly Trp His Ile Ala Ala Ala Gln Glu Ser  
 165 170 175  
 Ala Ala Arg Arg Asn Leu Leu Gly Met Ile Ser His Gln Cys Leu Tyr  
 180 185 190  
 Asn Leu Ala Val Arg His Pro Glu Leu Asp Val Leu Pro Ala Ala Gln  
 195 200 205  
 Ala Tyr Gly Val Gly Val Phe Ala Trp Ser Pro Leu His Gly Gly Leu  
 210 215 220  
 Leu Ser Gly Val Leu Glu Lys Leu Ala Ala Gly Thr Ala Val Lys Ser  
 225 230 235 240  
 Ala Gln Gly Arg Ala Gln Val Leu Leu Pro Ala Val Arg Pro Leu Val  
 245 250 255  
 Glu Ala Tyr Glu Asp Tyr Cys Arg Arg Leu Gly Ala Asp Pro Ala Glu  
 260 265 270  
 Val Gly Leu Ala Trp Val Leu Ser Arg Pro Gly Ile Leu Gly Ala Val  
 275 280 285  
 Ile Gly Pro Arg Thr Pro Glu Gln Leu Asp Ser Ala Leu Arg Ala Ala  
 290 295 300  
 Glu Leu Thr Leu Gly Glu Glu Glu Leu Arg Glu Leu Glu Ala Ile Phe  
 305 310 315 320  
 Pro Ala Pro Ala Val Asp Gly Pro Val Pro  
 325 330

<210> 11  
 <211> 417  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 11

Met Arg Val Leu Leu Thr Ser Phe Ala His Arg Thr His Phe Gln Gly  
 1 5 10 15  
 Leu Val Pro Leu Ala Trp Ala Leu His Thr Ala Gly His Asp Val Arg  
 20 25 30  
 Val Ala Ser Gln Pro Glu Leu Thr Asp Val Val Val Gly Ala Gly Leu  
 35 40 45  
 Thr Ser Val Pro Leu Gly Ser Asp His Arg Leu Phe Asp Ile Ser Pro  
 50 55 60  
 Glu Ala Ala Ala Gln Val His Arg Tyr Thr Thr Asp Leu Asp Phe Ala  
 65 70 75 80  
 Arg Arg Gly Pro Glu Leu Arg Ser Trp Glu Phe Leu His Gly Ile Glu  
 85 90 95  
 Glu Ala Thr Ser Arg Phe Val Phe Pro Val Val Asn Asn Asp Ser Phe  
 100 105 110  
 Val Asp Glu Leu Val Glu Phe Ala Met Asp Trp Arg Pro Asp Leu Val  
 115 120 125  
 Leu Trp Glu Pro Phe Thr Phe Ala Gly Ala Val Ala Ala Lys Ala Cys  
 130 135 140  
 Gly Ala Ala His Ala Arg Leu Leu Trp Gly Ser Asp Leu Thr Gly Tyr  
 145 150 155 160  
 Phe Arg Ser Arg Ser Gln Asp Leu Arg Gly Gln Arg Pro Ala Asp Asp  
 165 170 175  
 Arg Pro Asp Pro Leu Gly Gly Trp Leu Thr Glu Val Ala Gly Arg Phe  
 180 185 190  
 Gly Leu Asp Tyr Ser Glu Asp Leu Ala Val Gly Gln Trp Ser Val Asp  
 195 200 205  
 Gln Leu Pro Glu Ser Phe Arg Leu Glu Thr Gly Leu Glu Ser Val His  
 210 215 220  
 Thr Arg Thr Leu Pro Tyr Asn Gly Ser Ser Val Val Pro Gln Trp Leu  
 225 230 235 240  
 Arg Thr Ser Asp Gly Val Arg Arg Val Cys Phe Thr Gly Gly Tyr Ser  
 245 250 255  
 Ala Leu Gly Ile Thr Ser Asn Pro Gln Glu Phe Leu Arg Thr Leu Ala  
 260 265 270  
 Thr Leu Ala Arg Phe Asp Gly Glu Ile Val Val Thr Arg Ser Gly Leu  
 275 280 285  
 Asp Pro Ala Ser Val Pro Asp Asn Val Arg Leu Val Asp Phe Val Pro  
 290 295 300  
 Met Asn Ile Leu Leu Pro Gly Cys Ala Ala Val Ile His His Gly Gly  
 305 310 315 320  
 Ala Gly Ser Trp Ala Thr Ala Leu His His Gly Val Pro Gln Ile Ser  
 325 330 335  
 Val Ala His Glu Trp Asp Cys Val Leu Arg Gly Gln Arg Thr Ala Glu  
 340 345 350  
 Leu Gly Ala Gly Val Phe Leu Arg Pro Asp Glu Val Asp Ala Asp Thr  
 355 360 365  
 Leu Trp Gln Ala Leu Ala Thr Val Val Glu Asp Arg Ser His Ala Glu  
 370 375 380  
 Asn Ala Glu Lys Leu Arg Gln Glu Ala Leu Ala Ala Pro Thr Pro Ala  
 385 390 395 400  
 Glu Val Val Pro Val Leu Glu Ala Leu Ala His Gln His Arg Ala Asp  
 405 410 415  
 Arg

&lt;210&gt; 12

&lt;211&gt; 313

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 12

Met Thr Arg His Val Thr Leu Leu Gly Val Ser Gly Phe Val Gly Ser  
 1 5 10 15  
 Ala Leu Leu Arg Glu Phe Thr Thr His Pro Leu Arg Leu Arg Ala Val  
 20 25 30  
 Ala Arg Thr Gly Ser Arg Asp Gln Pro Pro Gly Ser Ala Gly Ile Glu  
 35 40 45  
 His Leu Arg Val Asp Leu Leu Glu Pro Gly Arg Val Ala Gln Val Val  
 50 55 60  
 Ala Asp Thr Asp Val Val Val His Leu Val Ala Tyr Ala Ala Gly Gly  
 65 70 75 80  
 Ser Thr Trp Arg Ser Ala Ala Thr Val Pro Glu Ala Glu Arg Val Asn  
 85 90 95  
 Ala Gly Ile Met Arg Asp Leu Val Ala Ala Leu Arg Ala Arg Pro Gly  
 100 105 110  
 Pro Ala Pro Val Leu Leu Phe Ala Ser Thr Thr Gln Ala Ala Asn Pro  
 115 120 125  
 Ala Ala Pro Ser Arg Tyr Ala Gln His Lys Ile Glu Ala Glu Arg Ile  
 130 135 140  
 Leu Arg Gln Ala Thr Glu Asp Gly Val Val Asp Gly Val Ile Leu Arg  
 145 150 155 160  
 Leu Pro Ala Ile Tyr Gly His Ser Gly Pro Ser Gly Gln Thr Gly Arg  
 165 170 175  
 Gly Val Val Thr Ala Met Ile Arg Arg Ala Leu Ala Gly Glu Pro Ile  
 180 185 190  
 Thr Met Trp His Glu Gly Ser Val Arg Arg Asn Leu Leu His Val Glu  
 195 200 205  
 Asp Val Ala Thr Ala Phe Thr Ala Ala Leu His Asn His Glu Ala Leu  
 210 215 220  
 Val Gly Asp Val Trp Thr Pro Ser Ala Asp Glu Ala Arg Pro Leu Gly  
 225 230 235 240  
 Glu Ile Phe Glu Thr Val Ala Ala Ser Val Ala Arg Gln Thr Gly Asn  
 245 250 255  
 Pro Ala Val Pro Val Val Ser Val Pro Pro Pro Glu Asn Ala Glu Ala  
 260 265 270  
 Asn Asp Phe Arg Ser Asp Asp Phe Asp Ser Thr Glu Phe Arg Thr Leu  
 275 280 285  
 Thr Gly Trp His Pro Arg Val Pro Leu Ala Glu Gly Ile Asp Arg Thr  
 290 295 300  
 Val Ala Ala Leu Ile Ser Thr Lys Glu  
 305 310

&lt;210&gt; 13

&lt;211&gt; 3546

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 13

Met Val Asp Val Pro Asp Leu Leu Gly Thr Arg Thr Pro His Pro Gly  
 1 5 10 15  
 Pro Leu Pro Phe Pro Trp Pro Leu Cys Gly His Asn Glu Pro Glu Leu  
 20 25 30  
 Arg Ala Arg Ala Arg Gln Leu His Ala Tyr Leu Glu Gly Ile Ser Glu  
 35 40 45  
 Asp Asp Val Val Ala Val Gly Ala Ala Leu Ala Arg Glu Thr Arg Ala  
 50 55 60  
 Gln Asp Gly Pro His Arg Ala Val Val Val Ala Ser Ser Val Thr Glu  
 65 70 75 80  
 Leu Thr Ala Ala Leu Ala Ala Leu Ala Gln Gly Arg Pro His Pro Ser  
 85 90 95  
 Val Val Arg Gly Val Ala Arg Pro Thr Ala Pro Val Val Phe Val Leu  
 100 105 110

Pro Gly Gln Gly Ala Gln Trp Pro Gly Met Ala Thr Arg Leu Leu Ala  
 115 120 125  
 Glu Ser Pro Val Phe Ala Ala Ala Met Arg Ala Cys Glu Arg Ala Phe  
 130 135 140  
 Asp Glu Val Thr Asp Trp Ser Leu Thr Glu Val Leu Asp Ser Pro Glu  
 145 150 155 160  
 His Leu Arg Arg Val Glu Val Val Gln Pro Ala Leu Phe Ala Val Gln  
 165 170 175  
 Thr Ser Leu Ala Ala Leu Trp Arg Ser Phe Gly Val Arg Pro Asp Ala  
 180 185 190  
 Val Leu Gly His Ser Ile Gly Glu Leu Ala Ala Ala Glu Val Cys Gly  
 195 200 205  
 Ala Val Asp Val Glu Ala Ala Ala Arg Ala Ala Ala Leu Trp Ser Arg  
 210 215 220  
 Glu Met Val Pro Leu Val Gly Arg Gly Asp Met Ala Ala Val Ala Leu  
 225 230 235 240  
 Ser Pro Ala Glu Leu Ala Ala Arg Val Glu Arg Trp Asp Asp Asp Val  
 245 250 255  
 Val Pro Ala Gly Val Asn Gly Pro Arg Ser Val Leu Leu Thr Gly Ala  
 260 265 270  
 Pro Glu Pro Ile Ala Arg Arg Val Ala Glu Leu Ala Ala Gln Gly Val  
 275 280 285  
 Arg Ala Gln Val Val Asn Val Ser Met Ala Ala His Ser Ala Gln Val  
 290 295 300  
 Asp Ala Val Ala Glu Gly Met Arg Ser Ala Leu Thr Trp Phe Ala Pro  
 305 310 315 320  
 Gly Asp Ser Asp Val Pro Tyr Tyr Ala Gly Leu Thr Gly Gly Arg Leu  
 325 330 335  
 Asp Thr Arg Glu Leu Gly Ala Asp His Trp Pro Arg Ser Phe Arg Leu  
 340 345 350  
 Pro Val Arg Phe Asp Glu Ala Thr Arg Ala Val Leu Glu Leu Gln Pro  
 355 360 365  
 Gly Thr Phe Ile Glu Ser Ser Pro His Pro Val Leu Ala Ala Ser Leu  
 370 375 380  
 Gln Gln Thr Leu Asp Glu Val Gly Ser Pro Ala Ala Ile Val Pro Thr  
 385 390 395 400  
 Leu Gln Arg Asp Gln Gly Gly Leu Arg Arg Phe Leu Leu Ala Val Ala  
 405 410 415  
 Gln Ala Tyr Thr Gly Gly Val Thr Val Asp Trp Thr Ala Ala Tyr Pro  
 420 425 430  
 Gly Val Thr Pro Gly His Leu Pro Ser Ala Val Ala Val Glu Thr Asp  
 435 440 445  
 Glu Gly Pro Ser Thr Glu Phe Asp Trp Ala Ala Pro Asp His Val Leu  
 450 455 460  
 Arg Ala Arg Leu Leu Glu Ile Val Gly Ala Glu Thr Ala Ala Leu Ala  
 465 470 475 480  
 Gly Arg Glu Val Asp Ala Arg Ala Thr Phe Arg Glu Leu Gly Leu Asp  
 485 490 495  
 Ser Val Leu Ala Val Gln Leu Arg Thr Arg Leu Ala Thr Ala Thr Gly  
 500 505 510  
 Arg Asp Leu His Ile Ala Met Leu Tyr Asp His Pro Thr Pro His Ala  
 515 520 525  
 Leu Thr Glu Ala Leu Leu Arg Gly Pro Gln Glu Glu Pro Gly Arg Gly  
 530 535 540  
 Glu Glu Thr Ala His Pro Thr Glu Ala Glu Pro Asp Glu Pro Val Ala  
 545 550 555 560  
 Val Val Ala Met Ala Cys Arg Leu Pro Gly Gly Val Thr Ser Pro Glu  
 565 570 575  
 Glu Phe Trp Glu Leu Leu Ala Glu Gly Arg Asp Ala Val Gly Gly Leu  
 580 585 590  
 Pro Thr Asp Arg Gly Trp Asp Leu Asp Ser Leu Phe His Pro Asp Pro

28

Gln Gly Trp Gln Trp Ala Gly Met Ala Val Asp Leu Leu Asp Gly Asp  
 1090 1095 1100  
 Pro Val Phe Ala Ser Val Leu Arg Glu Cys Ala Asp Ala Leu Glu Pro  
 1105 1110 1115 1120  
 Tyr Leu Asp Phe Glu Ile Val Pro Phe Leu Arg Ala Glu Ala Gln Arg  
 1125 1130 1135  
 Arg Thr Pro Asp His Thr Leu Ser Thr Asp Arg Val Asp Val Val Gln  
 1140 1145 1150  
 Pro Val Leu Phe Ala Val Met Val Ser Leu Ala Ala Arg Trp Arg Ala  
 1155 1160 1165  
 Tyr Gly Val Glu Pro Ala Ala Val Ile Gly His Ser Gln Gly Glu Ile  
 1170 1175 1180  
 Ala Ala Ala Cys Val Ala Gly Ala Leu Ser Leu Asp Asp Ala Ala Arg  
 1185 1190 1195 1200  
 Ala Val Ala Leu Arg Ser Arg Val Ile Ala Thr Met Pro Gly Asn Gly  
 1205 1210 1215  
 Ala Met Ala Ser Ile Ala Ala Ser Val Asp Glu Val Ala Ala Arg Ile  
 1220 1225 1230  
 Asp Gly Arg Val Glu Ile Ala Ala Val Asn Gly Pro Arg Ala Val Val  
 1235 1240 1245  
 Val Ser Gly Asp Arg Asp Asp Leu Asp Arg Leu Val Ala Ser Cys Thr  
 1250 1255 1260  
 Val Glu Gly Val Arg Ala Lys Arg Leu Pro Val Asp Tyr Ala Ser His  
 1265 1270 1275 1280  
 Ser Ser His Val Glu Ala Val Arg Asp Ala Leu His Ala Glu Leu Gly  
 1285 1290 1295  
 Glu Phe Arg Pro Leu Pro Gly Phe Val Pro Phe Tyr Ser Thr Val Thr  
 1300 1305 1310  
 Gly Arg Trp Val Glu Pro Ala Glu Leu Asp Ala Gly Tyr Trp Phe Arg  
 1315 1320 1325  
 Asn Leu Arg His Arg Val Arg Phe Ala Asp Ala Val Arg Ser Leu Ala  
 1330 1335 1340  
 Asp Gln Gly Tyr Thr Thr Phe Leu Glu Val Ser Ala His Pro Val Leu  
 1345 1350 1355 1360  
 Thr Thr Ala Ile Glu Glu Ile Gly Glu Asp Arg Gly Gly Asp Leu Val  
 1365 1370 1375  
 Ala Val His Ser Leu Arg Arg Gly Ala Gly Gly Pro Val Asp Phe Gly  
 1380 1385 1390  
 Ser Ala Leu Ala Arg Ala Phe Val Ala Gly Val Ala Val Asp Trp Glu  
 1395 1400 1405  
 Ser Ala Tyr Gln Gly Ala Gly Ala Arg Arg Val Pro Leu Pro Thr Tyr  
 1410 1415 1420  
 Pro Phe Gln Arg Glu Arg Phe Trp Leu Glu Pro Asn Pro Ala Arg Arg  
 1425 1430 1435 1440  
 Val Ala Asp Ser Asp Asp Val Ser Ser Leu Arg Tyr Arg Ile Glu Trp  
 1445 1450 1455  
 His Pro Thr Asp Pro Gly Glu Pro Gly Arg Leu Asp Gly Thr Trp Leu  
 1460 1465 1470  
 Leu Ala Thr Tyr Pro Gly Arg Ala Asp Asp Arg Val Glu Ala Ala Arg  
 1475 1480 1485  
 Gln Ala Leu Glu Ser Ala Gly Ala Arg Val Glu Asp Leu Val Val Glu  
 1490 1495 1500  
 Pro Arg Thr Gly Arg Val Asp Leu Val Arg Arg Leu Asp Ala Val Gly  
 1505 1510 1515 1520  
 Pro Val Ala Gly Val Leu Cys Leu Phe Ala Val Ala Glu Pro Ala Ala  
 1525 1530 1535  
 Glu His Ser Pro Leu Ala Val Thr Ser Leu Ser Asp Thr Leu Asp Leu  
 1540 1545 1550  
 Thr Gln Ala Val Ala Gly Ser Gly Arg Glu Cys Pro Ile Trp Val Val  
 1555 1560 1565  
 Thr Glu Asn Ala Val Ala Val Gly Pro Phe Glu Arg Leu Arg Asp Pro



1570	1575	1580
Ala His Gly Ala Leu Trp	Ala Leu Gly Arg Val	Val Ala Leu Glu Asn
1585	1590	1595
Pro Ala Val Trp Gly Gly Leu Val Asp	Val Pro Ser Gly Ser Val Ala	1600
1605	1610	1615
Glu Leu Ser Arg His Leu Gly Thr Thr	Leu Ser Gly Ala Gly Glu Asp	1620
1620	1625	1630
Gln Val Ala Leu Arg Pro Asp Gly Thr Tyr	Ala Arg Arg Trp Cys Arg	1635
1640	1645	1650
Ala Gly Ala Gly Gly Thr Gly Arg Trp Gln	Pro Arg Gly Thr Val Leu	1655
1660	1665	1670
Val Thr Gly Gly Thr Gly Gly Val Gly Arg	His Val Ala Arg Trp Leu	1675
1680	1685	1690
Ala Arg Gln Gly Thr Pro Cys Leu Val	Leu Ala Ser Arg Arg Gly Pro	1695
1700	1705	1710
Asp Ala Asp Gly Val Glu Glu Leu Leu Thr	Glu Leu Ala Asp Leu Gly	1715
1720	1725	1730
Thr Arg Ala Thr Val Thr Ala Cys Asp Val	Thr Asp Arg Glu Gln Leu	1735
1740	1745	1750
Arg Ala Leu Leu Ala Thr Val Asp Asp Gly	Thr Val Glu Thr Leu Thr	1755
1760	1765	1770
Gly Asp Arg Ile Glu Arg Ala Asn Arg Ala	Lys Val Leu Gly Ala Arg	1775
1780	1785	1790
Asn Leu His Glu Leu Thr Arg Asp Ala Asp	Leu Asp Ala Phe Val Leu	1795
1800	1805	1810
Phe Ser Ser Ser Thr Ala Ala Phe Gly Ala	Pro Gly Leu Gly Gly Tyr	1815
1820	1825	1830
Val Pro Gly Asn Ala Tyr Leu Asp Gly Leu	Ala Gln Gln Arg Arg Ser	1835
1840	1845	1850
Glu Gly Leu Pro Ala Thr Ser Val Ala Trp	Gly Thr Trp Ala Gly Ser	1855
1860	1865	1870
Gly Met Ala Glu Gly Pro Val Ala Asp Arg	Phe Arg Arg His Gly Val	1875
1880	1885	1890
Met Glu Met His Pro Asp Gln Ala Val Glu	Gly Leu Arg Val Ala Leu	1895
1900	1905	1910
Val Gln Gly Glu Val Ala Pro Ile Val Val	Asp Ile Arg Trp Asp Arg	1915
1920	1925	1930
Phe Leu Leu Ala Tyr Thr Ala Gln Arg Pro	Thr Arg Leu Phe Asp Thr	1935
1940	1945	1950
Ileu Asp Glu Ala Arg Arg Ala Ala Pro Gly	Pro Asp Ala Gly Pro Gly	1955
1960	1965	1970
Val Ala Ala Leu Ala Gly Leu Pro Val Gly	Glu Arg Glu Lys Ala Val	1975
1980	1985	1990
Leu Asp Leu Val Arg Thr His Ala Ala Val	Leu Gly His Ala Ser	1995
2000	2005	2010
Ala Glu Gln Val Pro Val Asp Arg Ala Phe	Ala Glu Leu Gly Val Asp	2015
2020	2025	2030
Ser Leu Ser Ala Leu Glu Leu Arg Asn Arg	Leu Thr Thr Ala Thr Gly	2035
2040	2045	2050
Val Arg Leu Ala Thr Thr Thr Val Phe Asp	His Pro Asp Val Arg Thr	2055
2060	2065	2070
Leu Ala Gly His Leu Ala Ala Glu Leu Gly	Gly Gly Gly Ser Gly Arg Glu	2075
2080	2085	2090
Arg Pro Gly Gly Glu Ala Pro Thr Val Ala	Pro Thr Asp Glu Pro Ile	2095
2100	2105	2110
Ala Ile Val Gly Met Ala Cys Arg Leu Pro	Gly Gly Val Asp Ser Pro	2115
2120	2125	2130
Glu Gln-Leu Trp Glu Leu Ile Val Ser Gly	Arg Asp Thr Ala Ser Ala	2135
2140	2145	2150

Ala Pro Gly Asp Arg Ser Trp Asp Pro Ala Glu Leu Met Val Ser Asp  
 2065 2070 2075 2080  
 Thr Thr Gly Thr Arg Thr Ala Phe Gly Asn Phe Met Pro Gly Ala Gly  
 2085 2090 2095  
 Glu Phe Asp Ala Ala Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala  
 2100 2105 2110  
 Met Asp Pro Gln Gln Arg His Ala Leu Glu Thr Thr Trp Glu Ala Leu  
 2115 2120 2125  
 Glu Asn Ala Gly Ile Arg Pro Glu Ser Leu Arg Gly Thr Asp Thr Gly  
 2130 2135 2140  
 Val Phe Val Gly Met Ser His Gln Gly Tyr Ala Thr Gly Arg Pro Lys  
 2145 2150 2155 2160  
 Pro Glu Asp Glu Val Asp Gly Tyr Leu Leu Thr Gly Asn Thr Ala Ser  
 2165 2170 2175  
 Val Ala Ser Gly Arg Ile Ala Tyr Val Leu Gly Leu Glu Gly Pro Ala  
 2180 2185 2190  
 Ile Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Leu His Val  
 2195 2200 2205  
 Ala Ala Gly Ser Leu Arg Ser Gly Asp Cys Gly Leu Ala Val Ala Gly  
 2210 2215 2220  
 Gly Val Ser Val Met Ala Gly Pro Glu Val Phe Arg Glu Phe Ser Arg  
 2225 2230 2235 2240  
 Gln Gly Ala Leu Ala Pro Asp Gly Arg Cys Lys Pro Phe Ser Asp Glu  
 2245 2250 2255  
 Ala Asp Gly Phe Gly Leu Gly Glu Gly Ser Ala Phe Val Val Leu Gln  
 2260 2265 2270  
 Arg Leu Ser Val Ala Val Arg Glu Gly Arg Arg Val Leu Gly Val Val  
 2275 2280 2285  
 Val Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Ala Ala  
 2290 2295 2300  
 Pro Ser Gly Val Ala Gln Gln Arg Val Ile Arg Arg Ala Trp Gly Arg  
 2305 2310 2315 2320  
 Ala Gly Val Ser Gly Gly Asp Val Gly Val Val Glu Ala His Gly Thr  
 2325 2330 2335  
 Gly Thr Arg Leu Gly Asp Pro Val Glu Leu Gly Ala Leu Leu Gly Thr  
 2340 2345 2350  
 Tyr Gly Val Gly Arg Gly Gly Val Gly Pro Val Val Val Gly Ser Val  
 2355 2360 2365  
 Lys Ala Asn Val Gly His Val Gln Ala Ala Ala Gly Val Val Gly Val  
 2370 2375 2380  
 Ile Lys Val Val Leu Gly Leu Gly Arg Gly Leu Val Gly Pro Met Val  
 2385 2390 2395 2400  
 Cys Arg Gly Gly Leu Ser Gly Leu Val Asp Trp Ser Ser Gly Gly Leu  
 2405 2410 2415  
 Val Val Ala Asp Gly Val Arg Gly Trp Pro Val Gly Val Asp Gly Val  
 2420 2425 2430  
 Arg Arg Gly Gly Val Ser Ala Phe Gly Val Ser Gly Thr Asn Ala His  
 2435 2440 2445  
 Val Val Val Ala Glu Ala Pro Gly Ser Val Val Gly Ala Glu Arg Pro  
 2450 2455 2460  
 Val Glu Gly Ser Ser Arg Gly Leu Val Gly Val Val Gly Gly Val Val  
 2465 2470 2475 2480  
 Pro Val Val Leu Ser Ala Lys Thr Glu Thr Ala Leu His Ala Gln Ala  
 2485 2490 2495  
 Arg Arg Leu Ala Asp His Leu Glu Thr His Pro Asp Val Pro Met Thr  
 2500 2505 2510  
 Asp Val Val Trp Thr Leu Thr Gln Ala Arg Gln Arg Phe Asp Arg Arg  
 2515 2520 2525  
 Ala Val Leu Leu Ala Ala Asp Arg Thr Gln Ala Val Glu Arg Leu Arg  
 2530 2535 2540  
 Gly Leu Ala Gly Gly Glu Pro Gly Thr Gly Val Val Ser Gly Val Ala

2545	2550	2555	2560
Ser Gly Gly Gly Val Val Phe Val Phe Pro Gly Gln Gly Gly Gln Trp			
	2565	2570	2575
Val Gly Met Ala Arg Gly Leu Leu Ser Val Pro Val Phe Val Glu Ser			
	2580	2585	2590
Val Val Glu Cys Asp Ala Val Val Ser Ser Val Val Gly Phe Ser Val			
	2595	2600	2605
Leu Gly Val Leu Glu Gly Arg Ser Gly Ala Pro Ser Leu Asp Arg Val			
	2610	2615	2620
Asp Val Val Gln Pro Val Leu Phe Val Val Met Val Ser Leu Ala Arg			
2625	2630	2635	2640
Leu Trp Arg Trp Cys Gly Val Val Pro Ala Ala Val Val Gly His Ser			
	2645	2650	2655
Gln Gly Glu Ile Ala Ala Ala Val Val Ala Gly Val Leu Ser Val Gly			
	2660	2665	2670
Asp Gly Ala Arg Val Val Ala Leu Arg Ala Arg Ala Leu Arg Ala Leu			
	2675	2680	2685
Ala Gly His Gly Gly Met Ala Ser Val Arg Arg Gly Arg Asp Asp Val			
	2690	2695	2700
Gln Lys Leu Leu Asp Ser Gly Pro Trp Thr Gly Lys Leu Glu Ile Ala			
2705	2710	2715	2720
Ala Val Asn Gly Pro Asp Ala Val Val Val Ser Gly Asp Pro Arg Ala			
	2725	2730	2735
Val Thr Glu Leu Val Glu His Cys Asp Gly Ile Gly Val Arg Ala Arg			
	2740	2745	2750
Thr Ile Pro Val Asp Tyr Ala Ser His Ser Ala Gln Val Glu Ser Leu			
	2755	2760	2765
Arg Glu Glu Leu Leu Ser Val Leu Ala Gly Ile Glu Gly Arg Pro Ala			
	2770	2775	2780
Thr Val Pro Phe Tyr Ser Thr Leu Thr Gly Gly Phe Val Asp Gly Thr			
2785	2790	2795	2800
Glu Leu Asp Ala Asp Tyr Trp Tyr Arg Asn Leu Arg His Pro Val Arg			
	2805	2810	2815
Phe His Ala Ala Val Glu Ala Leu Ala Ala Arg Asp Leu Thr Thr Phe			
	2820	2825	2830
Val Glu Val Ser Pro His Pro Val Leu Ser Met Ala Val Gly Glu Thr			
	2835	2840	2845
Leu Ala Asp Val Glu Ser Ala Val Thr Val Gly Thr Leu Glu Arg Asp			
	2850	2855	2860
Thr Asp Asp Val Glu Arg Phe Leu Thr Ser Leu Ala Glu Ala His Val			
2865	2870	2875	2880
His Gly Val Pro Val Asp Trp Ala Ala Val Leu Gly Ser Gly Thr Leu			
	2885	2890	2895
Val Asp Leu Pro Thr Tyr Pro Phe Gln Gly Arg Arg Phe Trp Leu His			
	2900	2905	2910
Pro Asp Arg Gly Pro Arg Asp Asp Val Ala Asp Trp Phe His Arg Val			
	2915	2920	2925
Asp Trp Thr Ala Thr Ala Thr Asp Gly Ser Ala Arg Leu Asp Gly Arg			
	2930	2935	2940
Trp Leu Val Val Val Pro Glu Gly Tyr Thr Asp Asp Gly Trp Val Val			
2945	2950	2955	2960
Glu Val Arg Ala Ala Leu Ala Ala Gly Gly Ala Glu Pro Val Val Thr			
	2965	2970	2975
Thr Val Glu Glu Val Thr Asp Arg Val Gly Asp Ser Asp Ala Val Val			
	2980	2985	2990
Ser Met Leu Gly Leu Ala Asp Asp Gly Ala Ala Glu Thr Leu Ala Leu			
	2995	3000	3005
Leu Arg Arg Leu Asp Ala Gln Ala Ser Thr Thr Pro Leu Trp Val Val			
	3010	3015	3020
Thr Val Gly Ala Val Ala Pro Ala Gly Pro Val Gln Arg Pro Glu Gln			
3025	3030	3035	3040

Ala Thr Val Trp Gly Leu Ala Leu Val Ala Ser Leu Glu Arg Gly His  
 3045 3050 3055  
 Arg Trp Thr Gly Leu Leu Asp Leu Pro Gln Thr Pro Asp Pro Gln Leu  
 3060 3065 3070  
 Arg Pro Arg Leu Val Glu Ala Leu Ala Gly Ala Glu Asp Gln Val Ala  
 3075 3080 3085  
 Val Arg Ala Asp Ala Val His Ala Arg Arg Ile Val Pro Thr Pro Val  
 3090 3095 3100  
 Thr Gly Ala Gly Pro Tyr Thr Ala Pro Gly Gly Thr Ile Leu Val Thr  
 3105 3110 3115 3120  
 Gly Gly Thr Ala Gly Leu Gly Ala Val Thr Ala Arg Trp Leu Ala Glu  
 3125 3130 3135  
 Arg Gly Ala Glu His Leu Ala Leu Val Ser Arg Arg Gly Pro Gly Thr  
 3140 3145 3150  
 Ala Gly Val Asp Glu Val Val Arg Asp Leu Thr Gly Leu Gly Val Arg  
 3155 3160 3165  
 Val Ser Val His Ser Cys Asp Val Gly Asp Arg Glu Ser Val Gly Ala  
 3170 3175 3180  
 Leu Val Gln Glu Leu Thr Ala Ala Gly Asp Val Val Arg Gly Val Val  
 3185 3190 3195 3200  
 His Ala Ala Gly Leu Pro Gln Gln Val Pro Leu Thr Asp Met Asp Pro  
 3205 3210 3215  
 Ala Asp Leu Ala Asp Val Val Ala Val Lys Val Asp Gly Ala Val His  
 3220 3225 3230  
 Leu Ala Asp Leu Cys Pro Glu Ala Glu Leu Phe Leu Leu Phe Ser Ser  
 3235 3240 3245  
 Gly Ala Gly Val Trp Gly Ser Ala Arg Gln Gly Ala Tyr Ala Ala Gly  
 3250 3255 3260  
 Asn Ala Phe Leu Asp Ala Phe Ala Arg His Arg Arg Asp Arg Gly Leu  
 3265 3270 3275 3280  
 Pro Ala Thr Ser Val Ala Trp Gly Leu Trp Ala Ala Gly Gly Met Thr  
 3285 3290 3295  
 Gly Asp Gln Glu Ala Val Ser Phe Leu Arg Glu Arg Gly Val Arg Pro  
 3300 3305 3310  
 Met Ser Val Pro Arg Ala Leu Glu Ala Leu Glu Arg Val Leu Thr Ala  
 3315 3320 3325  
 Gly Glu Thr Ala Val Val Val Ala Asp Val Asp Trp Ala Ala Phe Ala  
 3330 3335 3340  
 Glu Ser Tyr Thr Ser Ala Arg Pro Arg Pro Leu Leu His Arg Leu Val  
 3345 3350 3355 3360  
 Thr Pro Ala Ala Ala Val Gly Glu Arg Asp Glu Pro Arg Glu Gln Thr  
 3365 3370 3375  
 Leu Arg Asp Arg Leu Ala Ala Leu Pro Arg Ala Glu Arg Ser Ala Glu  
 3380 3385 3390  
 Leu Val Arg Leu Val Arg Arg Asp Ala Ala Ala Val Leu Gly Ser Asp  
 3395 3400 3405  
 Ala Lys Ala Val Pro Ala Thr Pro Phe Lys Asp Leu Gly Phe Asp  
 3410 3415 3420  
 Ser Leu Ala Ala Val Arg Phe Arg Asn Arg Leu Ala Ala His Thr Gly  
 3425 3430 3435 3440  
 Leu Arg Leu Pro Ala Thr Leu Val Phe Glu His Pro Asn Ala Ala Ala  
 3445 3450 3455  
 Val Ala Asp Leu Leu His Asp Arg Leu Gly Glu Ala Gly Glu Pro Thr  
 3460 3465 3470  
 Pro Val Arg Ser Val Gly Ala Gly Leu Ala Ala Leu Glu Gln Ala Leu  
 3475 3480 3485  
 Pro Asp Ala Ser Asp Thr Glu Arg Val Glu Leu Val Glu Arg Leu Glu  
 3490 3495 3500  
 Arg Met Leu Ala Gly Leu Arg Pro Glu Ala Gly Ala Gly Ala Asp Ala  
 3505 3510 3515 3520  
 Pro Thr Ala Gly Asp Asp Leu Gly Glu Ala Gly Val Asp Glu Leu Leu

3525                      3530                      3535  
 Asp Ala Leu Glu Arg Glu Leu Asp Ala Arg  
 3540                      3545

<210> 14  
 <211> 3562  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 14  
 Met Thr Asp Asn Asp Lys Val Ala Glu Tyr Leu Arg Arg Ala Thr Leu  
 1                      5                      10                      15  
 Asp Leu Arg Ala Ala Arg Lys Arg Leu Arg Glu Leu Gln Ser Asp Pro  
 20                      25                      30  
 Ile Ala Val Val Gly Met Ala Cys Arg Leu Pro Gly Gly Val His Leu  
 35                      40                      45  
 Pro Gln His Leu Trp Asp Leu Leu Arg Gln Gly His Glu Thr Val Ser  
 50                      55                      60  
 Thr Phe Pro Thr Gly Arg Gly Trp Asp Leu Ala Gly Leu Phe His Pro  
 65                      70                      75                      80  
 Asp Pro Asp His Pro Gly Thr Ser Tyr Val Asp Arg Gly Gly Phe Leu  
 85                      90                      95  
 Asp Asp Val Ala Gly Phe Asp Ala Glu Phe Phe Gly Ile Ser Pro Arg  
 100                      105                      110  
 Glu Ala Thr Ala Met Asp Pro Gln Gln Arg Leu Leu Leu Glu Thr Ser  
 115                      120                      125  
 Trp Glu Leu Val Glu Ser Ala Gly Ile Asp Pro His Ser Leu Arg Gly  
 130                      135                      140  
 Thr Pro Thr Gly Val Phe Leu Gly Val Ala Arg Leu Gly Tyr Gly Glu  
 145                      150                      155                      160  
 Asn Gly Thr Glu Ala Gly Asp Ala Glu Gly Tyr Ser Val Thr Gly Val  
 165                      170                      175  
 Ala Pro Ala Val Ala Ser Gly Arg Ile Ser Tyr Ala Leu Gly Leu Glu  
 180                      185                      190  
 Gly Pro Ser Ile Ser Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala  
 195                      200                      205  
 Leu His Leu Ala Val Glu Ser Leu Arg Leu Gly Glu Ser Ser Leu Ala  
 210                      215                      220  
 Val Val Gly Gly Ala Ala Val Met Ala Thr Pro Gly Val Phe Val Asp  
 225                      230                      235                      240  
 Phe Ser Arg Gln Arg Ala Leu Ala Ala Asp Gly Arg Ser Lys Ala Phe  
 245                      250                      255  
 Gly Ala Ala Ala Asp Gly Phe Gly Phe Ser Glu Gly Val Ser Leu Val  
 260                      265                      270  
 Leu Leu Glu Arg Leu Ser Glu Ala Glu Ser Asn Gly His Glu Val Leu  
 275                      280                      285  
 Ala Val Ile Arg Gly Ser Ala Leu Asn Gln Asp Gly Ala Ser Asn Gly  
 290                      295                      300  
 Leu Ala Ala Pro Asn Gly Thr Ala Gln Arg Lys Val Ile Arg Gln Ala  
 305                      310                      315                      320  
 Leu Arg Asn Cys Gly Leu Thr Pro Ala Asp Val Asp Ala Val Glu Ala  
 325                      330                      335  
 His Gly Thr Gly Thr Thr Leu Gly Asp Pro Ile Glu Ala Asn Ala Leu  
 340                      345                      350  
 Leu Asp Thr Tyr Gly Arg Asp Arg Asp Pro Asp His Pro Leu Trp Leu  
 355                      360                      365  
 Gly Ser Val Lys Ser Asn Ile Gly His Thr Gln Ala Ala Ala Gly Val  
 370                      375                      380  
 Thr Gly Leu Leu Lys Met Val Leu Ala Leu Arg His Glu Glu Leu Pro  
 385                      390                      395                      400  
 Ala Thr Leu His Val Asp Glu Pro Thr Pro His Val Asp Trp Ser Ser

35



Tyr	Pro	Phe	Gln	Arg	Lys	Pro	Tyr	Trp	Leu	Arg	Ser	Ser	Ala	Pro	Ala	900	905	910
Pro	Ala	Ser	His	Asp	Leu	Ala	Tyr	Arg	Val	Ser	Trp	Thr	Pro	Ile	Thr	915	920	925
Pro	Pro	Gly	Asp	Gly	Val	Leu	Asp	Gly	Asp	Trp	Leu	Val	Val	His	Pro	930	935	940
Gly	Gly	Ser	Thr	Gly	Trp	Val	Asp	Gly	Leu	Ala	Ala	Ala	Ile	Thr	Ala	945	950	955
Gly	Gly	Gly	Arg	Val	Val	Ala	His	Pro	Val	Asp	Ser	Val	Thr	Ser	Arg	965	970	975
Thr	Gly	Leu	Ala	Glu	Ala	Leu	Ala	Arg	Arg	Asp	Gly	Thr	Phe	Arg	Gly	980	985	990
Val	Leu	Ser	Trp	Val	Ala	Thr	Asp	Glu	Arg	His	Val	Glu	Ala	Gly	Ala	995	1000	1005
Val	Ala	Leu	Leu	Thr	Leu	Ala	Gln	Ala	Leu	Gly	Asp	Ala	Gly	Ile	Asp	1010	1015	1020
Ala	Pro	Leu	Trp	Cys	Leu	Thr	Gln	Glu	Ala	Val	Arg	Thr	Pro	Val	Asp	1025	1030	1035
Gly	Asp	Leu	Ala	Arg	Pro	Ala	Gln	Ala	Ala	Leu	His	Gly	Phe	Ala	Gln	1045	1050	1055
Val	Ala	Arg	Leu	Glu	Leu	Ala	Arg	Arg	Phe	Gly	Gly	Val	Leu	Asp	Leu	1060	1065	1070
Pro	Ala	Thr	Val	Asp	Ala	Ala	Gly	Thr	Arg	Leu	Val	Ala	Ala	Val	Leu	1075	1080	1085
Ala	Gly	Gly	Gly	Glu	Asp	Val	Val	Ala	Val	Arg	Gly	Asp	Arg	Leu	Tyr	1090	1095	1100
Gly	Arg	Arg	Leu	Val	Arg	Ala	Thr	Leu	Pro	Pro	Pro	Gly	Gly	Gly	Phe	1105	1110	1115
Thr	Pro	His	Gly	Thr	Val	Leu	Val	Thr	Gly	Ala	Ala	Gly	Pro	Val	Gly	1125	1130	1135
Gly	Arg	Leu	Ala	Arg	Trp	Leu	Ala	Glu	Arg	Gly	Ala	Thr	Arg	Leu	Val	1140	1145	1150
Leu	Pro	Gly	Ala	His	Pro	Gly	Glu	Glu	Leu	Leu	Thr	Ala	Ile	Arg	Ala	1155	1160	1165
Ala	Gly	Ala	Thr	Ala	Val	Val	Cys	Glu	Pro	Glu	Ala	Glu	Ala	Leu	Arg	1170	1175	1180
Thr	Ala	Ile	Gly	Gly	Glu	Leu	Pro	Thr	Ala	Leu	Val	His	Ala	Glu	Thr	1185	1190	1195
Leu	Thr	Asn	Phe	Ala	Gly	Val	Ala	Asp	Ala	Asp	Pro	Glu	Asp	Phe	Ala	1205	1210	1215
Ala	Thr	Val	Ala	Ala	Lys	Thr	Ala	Leu	Pro	Thr	Val	Leu	Ala	Glu	Val	1220	1225	1230
Leu	Gly	Asp	His	Arg	Leu	Glu	Arg	Glu	Val	Tyr	Cys	Ser	Ser	Val	Ala	1235	1240	1245
Gly	Val	Trp	Gly	Gly	Val	Gly	Met	Ala	Ala	Tyr	Ala	Ala	Gly	Ser	Ala	1250	1255	1260
Tyr	Leu	Asp	Ala	Leu	Val	Glu	His	Arg	Arg	Ala	Arg	Gly	His	Ala	Ser	1265	1270	1275
Ala	Ser	Val	Ala	Trp	Thr	Pro	Trp	Ala	Leu	Pro	Gly	Ala	Val	Asp	Asp	1285	1290	1295
Gly	Arg	Leu	Arg	Glu	Arg	Gly	Leu	Arg	Ser	Leu	Asp	Val	Ala	Asp	Ala	1300	1305	1310
Leu	Gly	Thr	Trp	Glu	Arg	Leu	Leu	Arg	Ala	Gly	Ala	Val	Ser	Val	Ala	1315	1320	1325
Val	Ala	Asp	Val	Asp	Trp	Ser	Val	Phe	Thr	Glu	Gly	Phe	Ala	Ala	Ile	1330	1335	1340
Arg	Pro	Thr	Pro	Leu	Phe	Asp	Glu	Leu	Leu	Asp	Arg	Arg	Gly	Asp	Pro	1345	1350	1355
Asp	Gly	Ala	Pro	Val	Asp	Arg	Pro	Gly	Glu	Pro	Ala	Gly	Glu	Trp	Gly	1365	1370	1375
Arg	Arg	Ile	Ala	Ala	Leu	Ser	Pro	Gln	Glu	Gln	Arg	Glu	Thr	Leu	Leu			

1380	1385	1390
Thr Leu Val Gly Glu Thr Val Ala Glu Val Leu Gly His Glu Thr Gly		
1395	1400	1405
Thr Glu Ile Asn Thr Arg Arg Ala Phe Ser Glu Leu Gly Leu Asp Ser		
1410	1415	1420
Leu Gly Ser Met Ala Leu Arg Gln Arg Leu Ala Ala Arg Thr Gly Leu		
1425	1430	1435
Arg Met Pro Ala Ser Leu Val Phe Asp His Pro Thr Val Thr Ala Leu		
1445	1450	1455
Ala Arg Tyr Leu Arg Arg Leu Val Val Gly Asp Ser Asp Pro Thr Pro		
1460	1465	1470
Val Arg Val Phe Gly Pro Thr Asp Glu Ala Glu Pro Val Ala Val Val		
1475	1480	1485
Gly Ile Gly Cys Arg Phe Pro Gly Gly Ile Ala Thr Pro Glu Asp Leu		
1490	1495	1500
Trp Arg Val Val Ser Glu Gly Thr Ser Ile Thr Thr Gly Phe Pro Thr		
1505	1510	1515
Asp Arg Gly Trp Asp Leu Arg Arg Leu Tyr His Pro Asp Pro Asp His		
1525	1530	1535
Pro Gly Thr Ser Tyr Val Asp Arg Gly Gly Phe Leu Asp Gly Ala Pro		
1540	1545	1550
Asp Phe Asp Pro Gly Phe Phe Gly Ile Thr Pro Arg Glu Ala Leu Ala		
1555	1560	1565
Met Asp Pro Gln Gln Arg Leu Thr Leu Glu Ile Ala Trp Glu Ala Val		
1570	1575	1580
Glu Arg Ala Gly Ile Asp Pro Glu Thr Leu Leu Gly Ser Asp Thr Gly		
1585	1590	1595
Val Phe Val Gly Met Asn Gly Gln Ser Tyr Leu Gln Leu Leu Thr Gly		
1605	1610	1615
Glu Gly Asp Arg Leu Asn Gly Tyr Gln Gly Leu Gly Asn Ser Ala Ser		
1620	1625	1630
Val Leu Ser Gly Arg Val Ala Tyr Thr Phe Gly Trp Glu Gly Pro Ala		
1635	1640	1645
Leu Thr Val Asp Thr Ala Cys Ser Ser Ser Leu Val Ala Ile His Leu		
1650	1655	1660
Ala Met Gln Ser Leu Arg Arg Gly Glu Cys Ser Leu Ala Leu Ala Gly		
1665	1670	1675
Gly Val Thr Val Met Ala Asp Pro Tyr Thr Phe Val Asp Phe Ser Ala		
1685	1690	1695
Gln Arg Gly Leu Ala Ala Asp Gly Arg Cys Lys Ala Phe Ser Ala Gln		
1700	1705	1710
Ala Asp Gly Phe Ala Leu Ala Glu Gly Val Ala Ala Leu Val Leu Glu		
1715	1720	1725
Pro Leu Ser Lys Ala Arg Arg Asn Gly His Gln Val Leu Ala Val Leu		
1730	1735	1740
Arg Gly Ser Ala Val Asn Gln Asp Gly Ala Ser Asn Gly Leu Ala Ala		
1745	1750	1755
Pro Asn Gly Pro Ser Gln Glu Arg Val Ile Arg Gln Ala Leu Thr Ala		
1765	1770	1775
Ser Gly Leu Arg Pro Ala Asp Val Asp Met Val Glu Ala His Gly Thr		
1780	1785	1790
Gly Thr Glu Leu Gly Asp Pro Ile Glu Ala Gly Ala Leu Ile Ala Ala		
1795	1800	1805
Tyr Gly Arg Asp Arg Asp Arg Pro Leu Trp Leu Gly Ser Val Lys Thr		
1810	1815	1820
Asn Ile Gly His Thr Gln Ala Ala Ala Gly Ala Ala Gly Val Ile Lys		
1825	1830	1835
Ala Val Leu Ala Met Arg His Gly Val Leu Pro Arg Ser Leu His Ala		
1845	1850	1855
Asp Glu Leu Ser Pro His Ile Asp Trp Ala Asp Gly Lys Val Glu Val		
1860	1865	1870

Leu Arg Glu Ala Arg Gln Trp Pro Pro Gly Glu Arg Pro Arg Arg Ala  
 1875 1880 1885  
 Gly Val Ser Ser Phe Gly Val Ser Gly Thr Asn Ala His Val Ile Val  
 1890 1895 1900  
 Glu Glu Ala Pro Ala Glu Pro Asp Pro Glu Pro Val Pro Ala Ala Pro  
 1905 1910 1915 1920  
 Gly Gly Pro Leu Pro Phe Val Leu His Gly Arg Ser Val Gln Thr Val  
 1925 1930 1935  
 Arg Ser Gln Ala Arg Thr Leu Ala Glu His Leu Arg Thr Thr Gly His  
 1940 1945 1950  
 Arg Asp Leu Ala Asp Thr Ala Arg Thr Leu Ala Thr Gly Arg Ala Arg  
 1955 1960 1965  
 Phe Asp Val Arg Ala Ala Val Leu Gly Thr Asp Arg Glu Gly Val Cys  
 1970 1975 1980  
 Ala Ala Leu Asp Ala Leu Ala Gln Asp Arg Pro Ser Pro Asp Val Val  
 1985 1990 1995 2000  
 Ala Pro Ala Val Phe Ala Ala Arg Thr Pro Val Leu Val Phe Pro Gly  
 2005 2010 2015  
 Gln Gly Ser Gln Trp Val Gly Met Ala Arg Asp Leu Leu Asp Ser Ser  
 2020 2025 2030  
 Glu Val Phe Ala Glu Ser Met Gly Arg Cys Ala Glu Ala Leu Ser Pro  
 2035 2040 2045  
 Tyr Thr Asp Trp Asp Leu Leu Asp Val Val Arg Gly Val Gly Asp Pro  
 2050 2055 2060  
 Asp Pro Tyr Asp Arg Val Asp Val Leu Gln Pro Val Leu Phe Ala Val  
 2065 2070 2075 2080  
 Met Val Ser Leu Ala Arg Leu Trp Gln Ser Tyr Gly Val Thr Pro Gly  
 2085 2090 2095  
 Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala His Val Ala  
 2100 2105 2110  
 Gly Ala Leu Ser Leu Ala Asp Ala Ala Arg Val Val Ala Leu Arg Ser  
 2115 2120 2125  
 Arg Val Leu Arg Glu Leu Asp Asp Gln Gly Gly Met Val Ser Val Gly  
 2130 2135 2140  
 Thr Ser Arg Ala Glu Leu Asp Ser Val Leu Arg Arg Trp Asp Gly Arg  
 2145 2150 2155 2160  
 Val Ala Val Ala Ala Val Asn Gly Pro Gly Thr Leu Val Val Ala Gly  
 2165 2170 2175  
 Pro Thr Ala Glu Leu Asp Glu Phe Leu Ala Val Ala Glu Ala Arg Glu  
 2180 2185 2190  
 Met Arg Pro Arg Arg Ile Ala Val Arg Tyr Ala Ser His Ser Pro Glu  
 2195 2200 2205  
 Val Ala Arg Val Glu Gln Arg Leu Ala Ala Glu Leu Gly Thr Val Thr  
 2210 2215 2220  
 Ala Val Gly Gly Thr Val Pro Leu Tyr Ser Thr Ala Thr Gly Asp Leu  
 2225 2230 2235 2240  
 Leu Asp Thr Thr Ala Met Asp Ala Gly Tyr Trp Tyr Arg Asn Leu Arg  
 2245 2250 2255  
 Gln Pro Val Leu Phe Glu His Ala Val Arg Ser Leu Leu Glu Arg Gly  
 2260 2265 2270  
 Phe Glu Thr Phe Ile Glu Val Ser Pro His Pro Val Leu Leu Met Ala  
 2275 2280 2285  
 Val Glu Glu Thr Ala Glu Asp Ala Glu Arg Pro Val Thr Gly Val Pro  
 2290 2295 2300  
 Thr Leu Arg Arg Asp His Asp Gly Pro Ser Glu Phe Leu Arg Asn Leu  
 2305 2310 2315 2320  
 Leu Gly Ala His Val His Gly Val Asp Val Asp Leu Arg Pro Ala Val  
 2325 2330 2335  
 Ala His Gly Arg Leu Val Asp Leu Pro Thr Tyr Pro Phe Asp Arg Gln  
 2340 2345 2350  
 Arg Leu Trp Pro Lys Pro His Arg Arg Ala Asp Thr Ser Ser Leu Gly

2355	2360	2365
Val Arg Asp Ser Thr His Pro Leu Leu His Ala Ala Val Asp Val Pro		
2370	2375	2380
Gly His Gly Gly Ala Val Phe Thr Gly Arg Leu Ser Pro Asp Glu Gln		
2385	2390	2395
Gln Trp Leu Thr Gln His Val Val Gly Gly Arg Asn Leu Val Pro Gly		2400
	2405	2410
Ser Val Leu Val Asp Leu Ala Leu Thr Ala Gly Ala Asp Val Gly Val		2415
	2420	2425
Pro Val Leu Glu Glu Leu Val Leu Gln Gln Pro Leu Val Leu Thr Ala		2430
	2435	2440
Ala Gly Ala Leu Leu Arg Leu Ser Val Gly Ala Ala Asp Glu Asp Gly		2445
	2450	2455
Arg Arg Pro Val Glu Ile His Ala Ala Glu Asp Val Ser Asp Pro Ala		2460
2465	2470	2475
Glu Ala Arg Trp Ser Ala Tyr Ala Thr Gly Thr Leu Ala Val Gly Val		2480
	2485	2490
Ala Gly Gly Gly Arg Asp Gly Thr Gln Trp Pro Pro Pro Gly Ala Thr		2495
	2500	2505
Ala Leu Thr Leu Thr Asp His Tyr Asp Thr Leu Ala Glu Leu Gly Tyr		2510
	2515	2520
Glu Tyr Gly Pro Ala Phe Gln Ala Leu Arg Ala Ala Trp Gln His Gly		2525
	2530	2535
Asp Val Val Tyr Ala Glu Val Ser Leu Asp Ala Val Glu Glu Gly Tyr		2540
2545	2550	2555
Ala Phe Asp Pro Val Leu Leu Asp Ala Val Ala Gln Thr Phe Gly Leu		2560
	2565	2570
Thr Ser Arg Ala Pro Gly Lys Leu Pro Phe Ala Trp Arg Gly Val Thr		2575
	2580	2585
Leu His Ala Thr Gly Ala Thr Ala Val Arg Val Val Ala Thr Pro Ala		2590
	2595	2600
Gly Pro Asp Ala Val Ala Leu Arg Val Thr Asp Pro Thr Gly Gln Leu		2605
	2610	2615
Val Ala Thr Val Asp Ala Leu Val Val Arg Asp Ala Gly Ala Asp Arg		2620
2625	2630	2635
Asp Gln Pro Arg Gly Arg Asp Gly Asp Leu His Arg Leu Glu Trp Val		2640
	2645	2650
Arg Leu Ala Thr Pro Asp Pro Thr Pro Ala Ala Val Val His Val Ala		2655
	2660	2665
Ala Asp Gly Leu Asp Asp Leu Leu Arg Ala Gly Gly Pro Ala Pro Gln		2670
	2675	2680
Ala Val Val Val Arg Tyr Arg Pro Asp Gly Asp Asp Pro Thr Ala Glu		2685
	2690	2695
Ala Arg His Gly Val Leu Trp Ala Ala Thr Leu Val Arg Arg Trp Leu		2700
2705	2710	2715
Asp Asp Asp Arg Trp Pro Ala Thr Thr Leu Val Val Ala Thr Ser Ala		2720
	2725	2730
Gly Val Glu Val Ser Pro Gly Asp Asp Val Pro Arg Pro Gly Ala Ala		2735
	2740	2745
Ala Val Trp Gly Val Leu Arg Cys Ala Gln Ala Glu Ser Pro Asp Arg		2750
	2755	2760
Phe Val Leu Val Asp Gly Asp Pro Glu Thr Pro Pro Ala Val Pro Asp		2765
	2770	2775
Asn Pro Gln Leu Ala Val Arg Asp Gly Ala Val Phe Val Pro Arg Leu		2780
2785	2790	2795
Thr Pro Leu Ala Gly Pro Val Pro Ala Val Ala Asp Arg Ala Tyr Arg		2800
	2805	2810
Leu Val Pro Gly Asn Gly Gly Ser Ile Glu Ala Val Ala Phe Ala Pro		2815
	2820	2825
Val Pro Asp Ala Asp Arg Pro Leu Ala Pro Glu Glu Val Arg Val Ala		2830
	2835	2840
		2845

Val Arg Ala Thr Gly Val Asn Phe Arg Asp Val Leu Leu Ala Leu Gly  
 2850 2855 2860  
 Met Tyr Pro Glu Pro Ala Glu Met Gly Thr Glu Ala Ser Gly Val Val  
 2865 2870 2875 2880  
 Thr Glu Val Gly Ser Gly Val Arg Arg Phe Thr Pro Gly Gln Ala Val  
 2885 2890 2895  
 Thr Gly Leu Phe Gln Gly Ala Phe Gly Pro Val Ala Val Ala Asp His  
 2900 2905 2910  
 Arg Leu Leu Thr Pro Val Pro Asp Gly Trp Arg Ala Val Asp Ala Ala  
 2915 2920 2925  
 Ala Val Pro Ile Ala Phe Thr Thr Ala His Tyr Ala Leu His Asp Leu  
 2930 2935 2940  
 Ala Gly Leu Gln Ala Gly Gln Ser Val Leu Val His Ala Ala Ala Gly  
 2945 2950 2955 2960  
 Gly Val Gly Met Ala Ala Val Ala Leu Ala Arg Arg Ala Gly Ala Glu  
 2965 2970 2975  
 Val Phe Ala Thr Ala Ser Pro Ala Lys His Pro Thr Leu Arg Ala Leu  
 2980 2985 2990  
 Gly Leu Asp Asp Asp His Ile Ala Ser Ser Arg Glu Ser Gly Phe Gly  
 2995 3000 3005  
 Glu Arg Phe Ala Ala Arg Thr Gly Gly Arg Gly Val Asp Val Val Leu  
 3010 3015 3020  
 Asn Ser Leu Thr Gly Asp Leu Leu Asp Glu Ser Ala Arg Leu Leu Ala  
 3025 3030 3035 3040  
 Asp Gly Gly Val Phe Val Glu Met Gly Lys Thr Asp Leu Arg Pro Ala  
 3045 3050 3055  
 Glu Gln Phe Arg Gly Arg Tyr Val Pro Phe Asp Leu Ala Glu Ala Gly  
 3060 3065 3070  
 Pro Asp Arg Leu Gly Glu Ile Leu Glu Glu Val Val Gly Leu Leu Ala  
 3075 3080 3085  
 Ala Gly Ala Leu Asp Arg Leu Pro Val Ser Val Trp Glu Leu Ser Ala  
 3090 3095 3100  
 Ala Pro Ala Ala Leu Thr His Met Ser Arg Gly Arg His Val Gly Lys  
 3105 3110 3115 3120  
 Leu Val Leu Thr Gln Pro Ala Pro Val His Pro Asp Gly Thr Val Leu  
 3125 3130 3135  
 Val Thr Gly Gly Thr Gly Thr Leu Gly Arg Leu Val Ala Arg His Leu  
 3140 3145 3150  
 Val Thr Gly His Gly Val Pro His Leu Leu Val Ala Ser Arg Arg Gly  
 3155 3160 3165  
 Pro Ala Ala Pro Gly Ala Ala Glu Leu Arg Ala Asp Val Glu Gly Leu  
 3170 3175 3180  
 Gly Ala Thr Ile Glu Ile Val Ala Cys Asp Thr Ala Asp Arg Glu Ala  
 3185 3190 3195 3200  
 Leu Ala Ala Leu Leu Asp Ser Ile Pro Ala Asp Arg Pro Leu Thr Gly  
 3205 3210 3215  
 Val Val His Thr Ala Gly Val Leu Ala Asp Gly Leu Val Thr Ser Ile  
 3220 3225 3230  
 Asp Gly Thr Ala Thr Asp Gln Val Leu Arg Ala Lys Val Asp Ala Ala  
 3235 3240 3245  
 Trp His Leu His Asp Leu Thr Arg Asp Ala Asp Leu Ser Phe Phe Val  
 3250 3255 3260  
 Leu Phe Ser Ser Ala Ala Ser Val Leu Ala Gly Pro Gly Gln Gly Val  
 3265 3270 3275 3280  
 Tyr Ala Ala Ala Asn Gly Val Leu Asn Ala Leu Ala Gly Gln Arg Arg  
 3285 3290 3295  
 Ala Leu Gly Leu Pro Ala Lys Ala Leu Gly Trp Gly Leu Trp Ala Gln  
 3300 3305 3310  
 Ala Ser Glu Met Thr Ser Gly Leu Gly Asp Arg Ile Ala Arg Thr Gly  
 3315 3320 3325  
 Val Ala Ala Leu Pro Thr Glu Arg Ala Leu Ala Leu Phe Asp Ala Ala

3330 3335 3340  
 Leu Arg Ser Gly Gly Glu Val Leu Phe Pro Leu Ser Val Asp Arg Ser  
 3345 3350 3355 3360  
 Ala Leu Arg Arg Ala Glu Tyr Val Pro Glu Val Leu Arg Gly Ala Val  
 3365 3370 3375  
 Arg Ser Thr Pro Arg Ala Ala Asn Arg Ala Glu Thr Pro Gly Arg Gly  
 3380 3385 3390  
 Leu Leu Asp Arg Leu Val Gly Ala Pro Glu Thr Asp Gln Val Ala Ala  
 3395 3400 3405  
 Leu Ala Glu Leu Val Arg Ser His Ala Ala Ala Val Ala Gly Tyr Asp  
 3410 3415 3420  
 Ser Ala Asp Gln Leu Pro Glu Arg Lys Ala Phe Lys Asp Leu Gly Phe  
 3425 3430 3435 3440  
 Asp Ser Leu Ala Ala Val Glu Leu Arg Asn Arg Leu Gly Val Thr Thr  
 3445 3450 3455  
 Gly Val Arg Leu Pro Ser Thr Leu Val Phe Asp His Pro Thr Pro Leu  
 3460 3465 3470  
 Ala Val Ala Glu His Leu Arg Ser Glu Leu Phe Ala Asp Ser Ala Pro  
 3475 3480 3485  
 Asp Val Gly Val Gly Ala Arg Leu Asp Asp Leu Glu Arg Ala Leu Asp  
 3490 3495 3500  
 Ala Leu Pro Asp Ala Gln Gly His Ala Asp Val Gly Ala Arg Leu Glu  
 3505 3510 3515 3520  
 Ala Leu Leu Arg Arg Trp Gln Ser Arg Arg Pro Pro Glu Thr Glu Pro  
 3525 3530 3535  
 Val Thr Ile Ser Asp Asp Ala Ser Asp Asp Glu Leu Phe Ser Met Leu  
 3540 3545 3550  
 Asp Arg Arg Leu Gly Gly Gly Gly Asp Val  
 3555 3560

&lt;210&gt; 15

&lt;211&gt; 3201

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 15

Met Ser Glu Ser Ser Gly Met Thr Glu Asp Arg Leu Arg Arg Tyr Leu  
 1 5 10 15  
 Lys Arg Thr Val Ala Glu Leu Asp Ser Val Thr Gly Arg Leu Asp Glu  
 20 25 30  
 Val Glu Tyr Arg Ala Arg Glu Pro Ile Ala Val Val Gly Met Ala Cys  
 35 40 45  
 Arg Phe Pro Gly Gly Val Asp Ser Pro Glu Ala Phe Trp Glu Phe Ile  
 50 55 60  
 Arg Asp Gly Gly Asp Ala Ile Ala Glu Ala Pro Thr Asp Arg Gly Trp  
 65 70 75 80  
 Pro Pro Ala Pro Arg Pro Arg Leu Gly Gly Leu Leu Ala Glu Pro Gly  
 85 90 95  
 Ala Phe Asp Ala Ala Phe Phe Gly Ile Ser Pro Arg Glu Ala Leu Ala  
 100 105 110  
 Thr Asp Pro Gln Gln Arg Leu Met Leu Glu Ile Ser Trp Glu Ala Leu  
 115 120 125  
 Glu Arg Ala Gly Phe Asp Pro Ser Ser Leu Arg Gly Ser Ala Gly Gly  
 130 135 140  
 Val Phe Thr Gly Val Gly Ala Val Asp Tyr Gly Pro Arg Pro Asp Glu  
 145 150 155 160  
 Ala Pro Glu Glu Val Leu Gly Tyr Val Gly Ile Gly Thr Ala Ser Ser  
 165 170 175  
 Val Ala Ser Gly Arg Val Ala Tyr Thr Leu Gly Leu Glu Gly Pro Ala  
 180 185 190  
 Val Thr Val Asp Thr Ala Cys Ser Ser Gly Leu Thr Ala Val His Leu



195	200	205
Ala Met Glu Ser Leu Arg Arg Asp Glu Cys Thr Leu Val Leu Ala Gly		
210	215	220
Gly Val Thr Val Met Ser Ser Pro Gly Ala Phe Thr Glu Phe Arg Ser		
225	230	235
Gln Gly Gly Leu Ala Glu Asp Gly Arg Cys Lys Pro Phe Ser Arg Ala		
245	250	255
Ala Asp Gly Phe Gly Leu Ala Glu Gly Ala Gly Val Leu Val Leu Gln		
260	265	270
Arg Leu Ser Val Ala Arg Ala Glu Gly Arg Pro Val Leu Ala Val Leu		
275	280	285
Arg Gly Ser Ala Ile Asn Gln Asp Gly Ala Ser Asn Gly Leu Thr Ala		
290	295	300
Pro Ser Gly Pro Ala Gln Arg Arg Val Ile Arg Gln Ala Leu Glu Arg		
305	310	315
Ala Arg Leu Arg Pro Val Asp Val Asp Tyr Val Glu Ala His Gly Thr		
325	330	335
Gly Thr Arg Leu Gly Asp Pro Ile Glu Ala His Ala Leu Leu Asp Thr		
340	345	350
Tyr Gly Ala Asp Arg Glu Pro Gly Arg Pro Leu Trp Val Gly Ser Val		
355	360	365
Lys Ser Asn Ile Gly His Thr Gln Ala Ala Ala Gly Val Ala Gly Val		
370	375	380
Met Lys Thr Val Leu Ala Leu Arg His Arg Glu Ile Pro Ala Thr Leu		
385	390	395
His Phe Asp Glu Pro Ser Pro His Val Asp Trp Asp Arg Gly Ala Val		
405	410	415
Ser Val Val Ser Glu Thr Arg Pro Trp Pro Val Gly Glu Arg Pro Arg		
420	425	430
Arg Ala Gly Val Ser Ser Phe Gly Ile Ser Gly Thr Asn Ala His Val		
435	440	445
Ile Val Glu Glu Ala Pro Ser Pro Gln Ala Ala Asp Leu Asp Pro Thr		
450	455	460
Pro Gly Pro Ala Thr Gly Ala Thr Pro Gly Thr Asp Ala Ala Pro Thr		
465	470	475
Ala Glu Pro Gly Ala Glu Ala Val Ala Leu Val Phe Ser Ala Arg Asp		
485	490	495
Glu Arg Ala Leu Arg Ala Gln Ala Ala Arg Leu Ala Asp Arg Leu Thr		
500	505	510
Asp Asp Pro Ala Pro Ser Leu Arg Asp Thr Ala Phe Thr Leu Val Thr		
515	520	525
Arg Arg Ala Thr Trp Glu His Arg Ala Val Val Val Gly Gly Gly Glu		
530	535	540
Glu Val Leu Ala Gly Leu Arg Ala Val Ala Gly Gly Arg Pro Val Asp		
545	550	555
Gly Ala Val Ser Gly Arg Ala Arg Ala Gly Arg Arg Val Val Leu Val		
565	570	575
Phe Pro Gly Gln Gly Ala Gln Trp Gln Gly Met Ala Arg Asp Leu Leu		
580	585	590
Arg Gln Ser Pro Thr Phe Ala Glu Ser Ile Asp Ala Cys Glu Arg Ala		
595	600	605
Leu Ala Pro His Val Asp Trp Ser Leu Arg Glu Val Leu Asp Gly Glu		
610	615	620
Gln Ser Leu Asp Pro Val Asp Val Val Gln Pro Val Leu Phe Ala Val		
625	630	635
Met Val Ser Leu Ala Arg Leu Trp Gln Ser Tyr Gly Val Thr Pro Gly		
645	650	655
Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala His Val Ala		
660	665	670
Gly Ala Leu Ser Leu Ala Asp Ala Ala Arg Val Val Ala Leu Arg Ser		
675	680	685

Arg Val Leu Arg Arg Leu Gly Gly His Gly Gly Met Ala Ser Phe Gly  
 690 695 700  
 Leu His Pro Asp Gln Ala Ala Glu Arg Ile Ala Arg Phe Ala Gly Ala  
 705 710 715 720  
 Leu Thr Val Ala Ser Val Asn Gly Pro Arg Ser Val Val Leu Ala Gly  
 725 730 735  
 Glu Asn Gly Pro Leu Asp Glu Leu Ile Ala Glu Cys Glu Ala Glu Gly  
 740 745 750  
 Val Thr Ala Arg Arg Ile Pro Val Asp Tyr Ala Ser His Ser Pro Gln  
 755 760 765  
 Val Glu Ser Leu Arg Glu Glu Leu Leu Ala Ala Leu Ala Gly Val Arg  
 770 775 780  
 Pro Val Ser Ala Gly Ile Pro Leu Tyr Ser Thr Leu Thr Gly Gln Val  
 785 790 795 800  
 Ile Glu Thr Ala Thr Met Asp Ala Asp Tyr Trp Phe Ala Asn Leu Arg  
 805 810 815  
 Glu Pro Val Arg Phe Gln Asp Ala Thr Arg Gln Leu Ala Glu Ala Gly  
 820 825 830  
 Phe Asp Ala Phe Val Glu Val Ser Pro His Pro Val Leu Thr Val Gly  
 835 840 845  
 Val Glu Ala Thr Leu Glu Ala Val Leu Pro Pro Asp Ala Asp Pro Cys  
 850 855 860  
 Val Thr Gly Thr Leu Arg Arg Glu Arg Gly Gly Leu Ala Gln Phe His  
 865 870 875 880  
 Thr Ala Leu Ala Glu Ala Tyr Thr Arg Gly Val Glu Val Asp Trp Arg  
 885 890 895  
 Thr Ala Val Gly Glu Gly Arg Pro Val Asp Leu Pro Val Tyr Pro Phe  
 900 905 910  
 Gln Arg Gln Asn Phe Trp Leu Pro Val Pro Leu Gly Arg Val Pro Asp  
 915 920 925  
 Thr Gly Asp Glu Trp Arg Tyr Gln Leu Ala Trp His Pro Val Asp Leu  
 930 935 940  
 Gly Arg Ser Ser Leu Ala Gly Arg Val Leu Val Val Thr Gly Ala Ala  
 945 950 955 960  
 Val Pro Pro Ala Trp Thr Asp Val Val Arg Asp Gly Leu Glu Gln Arg  
 965 970 975  
 Gly Ala Thr Val Val Leu Cys Thr Ala Gln Ser Arg Ala Arg Ile Gly  
 980 985 990  
 Ala Ala Leu Asp Ala Val Asp Gly Thr Ala Leu Ser Thr Val Val Ser  
 995 1000 1005  
 Leu Leu Ala Leu Ala Glu Gly Gly Ala Val Asp Asp Pro Ser Leu Asp  
 1010 1015 1020  
 Thr Leu Ala Leu Val Gln Ala Leu Gly Ala Ala Gly Ile Asp Val Pro  
 1025 1030 1035 1040  
 Leu Trp Leu Val Thr Arg Asp Ala Ala Ala Val Thr Val Gly Asp Asp  
 1045 1050 1055  
 Val Asp Pro Ala Gln Ala Met Val Gly Gly Leu Gly Arg Val Val Gly  
 1060 1065 1070  
 Val Glu Ser Pro Ala Arg Trp Gly Gly Leu Val Asp Leu Arg Glu Ala  
 1075 1080 1085  
 Asp Ala Asp Ser Ala Arg Ser Leu Ala Ala Ile Leu Ala Asp Pro Arg  
 1090 1095 1100  
 Gly Glu Glu Gln Phe Ala Ile Arg Pro Asp Gly Val Thr Val Ala Arg  
 1105 1110 1115 1120  
 Leu Val Pro Ala Pro Ala Arg Ala Ala Gly Thr Arg Trp Thr Pro Arg  
 1125 1130 1135  
 Gly Thr Val Leu Val Thr Gly Gly Thr Gly Gly Ile Gly Ala His Leu  
 1140 1145 1150  
 Ala Arg Trp Leu Ala Gly Ala Gly Ala Glu His Leu Val Leu Leu Asn  
 1155 1160 1165  
 Arg Arg Gly Ala Glu Ala Ala Gly Ala Ala Asp Leu Arg Asp Glu Leu

1170	1175	1180
Val Ala Leu Gly Thr Gly Val Thr Ile Thr Ala Cys Asp Val Ala Asp		
1185	1190	1195
Arg Asp Arg Leu Ala Val Leu Asp Ala Ala Arg Ala Gln Gly Arg		1200
	1205	1210
Val Val Thr Ala Val Phe His Ala Ala Gly Ile Ser Arg Ser Thr Ala		1215
	1220	1225
Val Gln Glu Leu Thr Glu Ser Glu Phe Thr Glu Ile Thr Asp Ala Lys		1230
	1235	1240
Val Arg Gly Thr Ala Asn Leu Ala Glu Leu Cys Pro Glu Leu Asp Ala		1245
	1250	1255
Leu Val Leu Phe Ser Ser Asn Ala Ala Val Trp Gly Ser Pro Gly Leu		1260
1265	1270	1275
Ala Ser Tyr Ala Ala Gly Asn Ala Phe Leu Asp Ala Phe Ala Arg Arg		1280
	1285	1290
Gly Arg Arg Ser Gly Leu Pro Val Thr Ser Ile Ala Trp Gly Leu Trp		1295
	1300	1305
Ala Gly Gln Asn Met Ala Gly Thr Glu Gly Gly Asp Tyr Leu Arg Ser		1310
	1315	1320
Gln Gly Leu Arg Ala Met Asp Pro Gln Arg Ala Ile Glu Glu Leu Arg		1325
	1330	1335
Thr Thr Leu Asp Ala Gly Asp Pro Trp Val Ser Val Val Asp Leu Asp		1340
1345	1350	1355
Arg Glu Arg Phe Val Glu Leu Phe Thr Ala Ala Arg Arg Arg Pro Leu		1360
	1365	1370
Phe Asp Glu Leu Gly Gly Val Arg Ala Gly Ala Glu Glu Thr Gly Gln		1375
	1380	1385
Glu Ser Asp Leu Ala Arg Arg Leu Ala Ser Met Pro Glu Ala Glu Arg		1390
	1395	1400
His Glu His Val Ala Arg Leu Val Arg Ala Glu Val Ala Ala Val Leu		1405
	1410	1415
Gly His Gly Thr Pro Thr Val Ile Glu Arg Asp Val Ala Phe Arg Asp		1420
1425	1430	1435
Leu Gly Phe Asp Ser Met Thr Ala Val Asp Leu Arg Asn Arg Leu Ala		1440
	1445	1450
Ala Val Thr Gly Val Arg Val Ala Thr Thr Ile Val Phe Asp His Pro		1455
	1460	1465
Thr Val Asp Arg Leu Thr Ala His Tyr Leu Glu Arg Leu Val Gly Glu		1470
	1475	1480
Pro Glu Ala Thr Thr Pro Ala Ala Ala Val Val Pro Gln Ala Pro Gly		1485
	1490	1495
Glu Ala Asp Glu Pro Ile Ala Ile Val Gly Met Ala Cys Arg Leu Ala		1500
1505	1510	1515
Gly Gly Val Arg Thr Pro Asp Gln Leu Trp Asp Phe Ile Val Ala Asp		1520
	1525	1530
Gly Asp Ala Val Thr Glu Met Pro Ser Asp Arg Ser Trp Asp Leu Asp		1535
	1540	1545
Ala Leu Phe Asp Pro Asp Pro Glu Arg His Gly Thr Ser Tyr Ser Arg		1550
	1555	1560
His Gly Ala Phe Leu Asp Gly Ala Ala Asp Phe Asp Ala Ala Phe Phe		1565
	1570	1575
Gly Ile Ser Pro Arg Glu Ala Leu Ala Met Asp Pro Gln Gln Arg Gln		1580
1585	1590	1595
Val Leu Glu Thr Thr Trp Glu Leu Phe Glu Asn Ala Gly Ile Asp Pro		1600
	1605	1610
His Ser Leu Arg Gly Thr Asp Thr Gly Val Phe Leu Gly Ala Ala Tyr		1615
	1620	1625
Gln Gly Tyr Gly Gln Asn Ala Gln Val Pro Lys Glu Ser Glu Gly Tyr		1630
	1635	1640
Leu Leu Thr Gly Gly Ser Ser Ala Val Ala Ser Gly Arg Ile Ala Tyr		1645
	1650	1655
		1660

Val Leu Gly Leu Glu Gly Pro Ala Ile Thr Val Asp Thr Ala Cys Ser  
 1665 1670 1675 1680  
 Ser Ser Leu Val Ala Leu His Val Ala Ala Gly Ser Leu Arg Ser Gly  
 1685 1690 1695  
 Asp Cys Gly Leu Ala Val Ala Gly Gly Val Ser Val Met Ala Gly Pro  
 1700 1705 1710  
 Glu Val Phe Thr Glu Phe Ser Arg Gln Gly Ala Leu Ala Pro Asp Gly  
 1715 1720 1725  
 Arg Cys Lys Pro Phe Ser Asp Gln Ala Asp Gly Phe Gly Phe Ala Glu  
 1730 1735 1740  
 Gly Val Ala Val Val Leu Leu Gln Arg Leu Ser Val Ala Val Arg Glu  
 1745 1750 1755 1760  
 Gly Arg Arg Val Leu Gly Val Val Val Gly Ser Ala Val Asn Gln Asp  
 1765 1770 1775  
 Gly Ala Ser Asn Gly Leu Ala Ala Pro Ser Gly Val Ala Gln Gln Arg  
 1780 1785 1790  
 Val Ile Arg Arg Ala Trp Gly Arg Ala Gly Val Ser Gly Gly Asp Val  
 1795 1800 1805  
 Gly Val Val Glu Ala His Gly Thr Gly Thr Arg Leu Gly Asp Pro Val  
 1810 1815 1820  
 Glu Leu Gly Ala Leu Leu Gly Thr Tyr Gly Val Gly Arg Gly Gly Val  
 1825 1830 1835 1840  
 Gly Pro Val Val Val Gly Ser Val Lys Ala Asn Val Gly His Val Gln  
 1845 1850 1855  
 Ala Ala Ala Gly Val Val Gly Val Ile Lys Val Val Leu Gly Leu Gly  
 1860 1865 1870  
 Arg Gly Leu Val Gly Pro Met Val Cys Arg Gly Gly Leu Ser Gly Leu  
 1875 1880 1885  
 Val Asp Trp Ser Ser Gly Gly Leu Val Val Ala Asp Gly Val Arg Gly  
 1890 1895 1900  
 Trp Pro Val Gly Val Asp Gly Val Arg Arg Gly Gly Val Ser Ala Phe  
 1905 1910 1915 1920  
 Gly Val Ser Gly Thr Asn Ala His Val Val Val Ala Glu Ala Pro Gly  
 1925 1930 1935  
 Ser Val Val Gly Ala Glu Arg Pro Val Glu Gly Ser Ser Arg Gly Leu  
 1940 1945 1950  
 Val Gly Val Ala Gly Gly Val Val Pro Val Val Leu Ser Ala Lys Thr  
 1955 1960 1965  
 Glu Thr Ala Leu Thr Glu Leu Ala Arg Arg Leu His Asp Ala Val Asp  
 1970 1975 1980  
 Asp Thr Val Ala Leu Pro Ala Val Ala Ala Thr Leu Ala Thr Gly Arg  
 1985 1990 1995 2000  
 Ala His Leu Pro Tyr Arg Ala Ala Leu Leu Ala Arg Asp His Asp Glu  
 2005 2010 2015  
 Leu Arg Asp Arg Leu Arg Ala Phe Thr Thr Gly Ser Ala Ala Pro Gly  
 2020 2025 2030  
 Val Val Ser Gly Val Ala Ser Gly Gly Gly Val Val Phe Val Phe Pro  
 2035 2040 2045  
 Gly Gln Gly Gly Gln Trp Val Gly Met Ala Arg Gly Leu Leu Ser Val  
 2050 2055 2060  
 Pro Val Phe Val Glu Ser Val Val Glu Cys Asp Ala Val Val Ser Ser  
 2065 2070 2075 2080  
 Val Val Gly Phe Ser Val Leu Gly Val Leu Glu Gly Arg Ser Gly Ala  
 2085 2090 2095  
 Pro Ser Leu Asp Arg Val Asp Val Val Gln Pro Val Leu Phe Val Val  
 2100 2105 2110  
 Met Val Ser Leu Ala Arg Leu Trp Arg Trp Cys Gly Val Val Pro Ala  
 2115 2120 2125  
 Ala Val Val Gly His Ser Gln Gly Glu Ile Ala Ala Ala Val Val Ala  
 2130 2135 2140  
 Gly Val Leu Ser Val Gly Asp Gly Ala Arg Val Val Ala Leu Arg Ala

2145	2150	2155	2160
Arg Ala Leu Arg	Ala Leu Ala Gly His Gly Gly Met Val Ser Leu Ala		
	2165	2170	2175
Val Ser Ala Glu Arg	Ala Arg Glu Leu Ile Ala Pro Trp Ser Asp Arg		
	2180	2185	2190
Ile Ser Val Ala Ala	Val Asn Ser Pro Thr Ser Val Val Val Ser Gly		
	2195	2200	2205
Asp Pro Gln Ala Leu	Ala Ala Leu Val Ala His Cys Ala Glu Thr Gly		
	2210	2215	2220
Glu Arg Ala Lys Thr	Leu Pro Val Asp Tyr Ala Ser His Ser Ala His		
	2225	2230	2235
Val Glu Gln Ile Arg	Asp Thr Ile Leu Thr Asp Leu Ala Asp Val Thr		
	2245	2250	2255
Ala Arg Arg Pro Asp	Val Ala Leu Tyr Ser Thr Leu His Gly Ala Arg		
	2260	2265	2270
Gly Ala Gly Thr Asp	Met Asp Ala Arg Tyr Trp Tyr Asp Asn Leu Arg		
	2275	2280	2285
Ser Pro Val Arg Phe	Asp Glu Ala Val Glu Ala Ala Val Ala Asp Gly		
	2290	2295	2300
Tyr Arg Val Phe Val	Glu Met Ser Pro His Pro Val Leu Thr Ala Ala		
	2305	2310	2315
Val Gln Glu Ile Asp	Asp Glu Thr Val Ala Ile Gly Ser Leu His Arg		
	2325	2330	2335
Asp Thr Gly Glu Arg	His Leu Val Ala Glu Leu Ala Arg Ala His Val		
	2340	2345	2350
His Gly Val Pro Val	Asp Trp Arg Ala Ile Leu Pro Ala Thr His Pro		
	2355	2360	2365
Val Pro Leu Pro Asn	Tyr Pro Phe Glu Ala Thr Arg Tyr Trp Leu Ala		
	2370	2375	2380
Pro Thr Ala Ala Asp	Gln Val Ala Asp His Arg Tyr Arg Val Asp Trp		
	2385	2390	2395
Arg Pro Leu Ala Thr	Thr Pro Ala Glu Leu Ser Gly Ser Tyr Leu Val		
	2405	2410	2415
Phe Gly Asp Ala Pro	Glu Thr Leu Gly His Ser Val Glu Lys Ala Gly		
	2420	2425	2430
Gly Leu Leu Val Pro	Val Ala Ala Pro Asp Arg Glu Ser Leu Ala Val		
	2435	2440	2445
Ala Leu Asp Glu Ala	Ala Gly Arg Leu Ala Gly Val Leu Ser Phe Ala		
	2450	2455	2460
Ala Asp Thr Ala Thr	His Leu Ala Arg His Arg Leu Leu Gly Glu Ala		
	2465	2470	2475
Asp Val Glu Ala Pro	Leu Trp Leu Val Thr Ser Gly Gly Val Ala Leu		
	2485	2490	2495
Asp Asp His Asp Pro	Ile Asp Cys Asp Gln Ala Met Val Trp Gly Ile		
	2500	2505	2510
Gly Arg Val Met Gly	Leu Glu Thr Pro His Arg Trp Gly Gly Leu Val		
	2515	2520	2525
Asp Val Thr Val Glu	Pro Thr Ala Glu Asp Gly Val Val Phe Ala Ala		
	2530	2535	2540
Leu Leu Ala Ala Asp	Asp His Glu Asp Gln Val Ala Leu Arg Asp Gly		
	2545	2550	2555
Ile Arg His Gly Arg	Arg Arg Leu Val Arg Ala Pro Leu Thr Thr Arg Asn		
	2565	2570	2575
Ala Arg Trp Thr Pro	Ala Gly Thr Ala Leu Val Thr Gly Gly Thr Gly		
	2580	2585	2590
Ala Leu Gly Gly His	Val Ala Arg Tyr Leu Ala Arg Ser Gly Val Thr		
	2595	2600	2605
Asp Leu Val Leu Leu	Ser Arg Ser Gly Pro Asp Ala Pro Gly Ala Ala		
	2610	2615	2620
Glu Leu Ala Ala Glu	Leu Ala Asp Leu Gly Ala Glu Pro Arg Val Glu		
	2625	2630	2635
			2640

Ala Cys Asp Val Thr Asp Gly Pro Arg Leu Arg Ala Leu Val Gln Glu  
 2645 2650 2655  
 Leu Arg Glu Gln Asp Arg Pro Val Arg Ile Val Val His Thr Ala Gly  
 2660 2665 2670  
 Val Pro Asp Ser Arg Pro Leu Asp Arg Ile Asp Glu Leu Glu Ser Val  
 2675 2680 2685  
 Ser Ala Ala Lys Val Thr Gly Ala Arg Leu Leu Asp Glu Leu Cys Pro  
 2690 2695 2700  
 Asp Ala Asp Thr Phe Val Leu Phe Ser Ser Gly Ala Gly Val Trp Gly  
 2705 2710 2715 2720  
 Ser Ala Asn Leu Gly Ala Tyr Ala Ala Ala Asn Ala Tyr Leu Asp Ala  
 2725 2730 2735  
 Leu Ala His Arg Arg Arg Gln Ala Gly Arg Ala Ala Thr Ser Val Ala  
 2740 2745 2750  
 Trp Gly Ala Trp Ala Gly Asp Gly Met Ala Thr Gly Asp Leu Asp Gly  
 2755 2760 2765  
 Leu Thr Arg Arg Gly Leu Arg Ala Met Ala Pro Asp Arg Ala Leu Arg  
 2770 2775 2780  
 Ala Cys Thr Arg Arg Trp Thr Thr His Asp Thr Cys Val Ser Val Ala  
 2785 2790 2795 2800  
 Asp Val Asp Trp Asp Arg Phe Ala Val Gly Phe Thr Ala Ala Arg Pro  
 2805 2810 2815  
 Arg Pro Leu Ile Asp Glu Leu Val Thr Ser Ala Pro Val Ala Ala Pro  
 2820 2825 2830  
 Thr Ala Ala Ala Ala Pro Val Pro Ala Met Thr Ala Asp Gln Leu Leu  
 2835 2840 2845  
 Gln Phe Thr Arg Ser His Val Ala Ala Ile Leu Gly His Gln Asp Pro  
 2850 2855 2860  
 Asp Ala Val Gly Leu Asp Gln Pro Phe Thr Glu Leu Gly Phe Asp Ser  
 2865 2870 2875 2880  
 Leu Thr Ala Val Gly Leu Arg Asn Gln Leu Gln Gln Ala Thr Gly Arg  
 2885 2890 2895  
 Thr Leu Pro Ala Ala Leu Val Phe Gln His Pro Thr Val Arg Arg Leu  
 2900 2905 2910  
 Ala Asp His Leu Ala Gln Gln Leu Asp Val Gly Thr Ala Pro Val Glu  
 2915 2920 2925  
 Ala Thr Gly Ser Val Leu Arg Asp Gly Tyr Arg Arg Ala Gly Gln Thr  
 2930 2935 2940  
 Gly Asp Val Arg Ser Tyr Leu Asp Leu Leu Ala Asn Leu Ser Glu Phe  
 2945 2950 2955 2960  
 Arg Glu Arg Phe Thr Asp Ala Ala Ser Leu Gly Gly Gln Leu Glu Leu  
 2965 2970 2975  
 Val Asp Leu Ala Asp Gly Ser Gly Pro Val Thr Val Ile Cys Cys Ala  
 2980 2985 2990  
 Gly Thr Ala Ala Leu Ser Gly Pro His Glu Phe Ala Arg Leu Ala Ser  
 2995 3000 3005  
 Ala Leu Arg Gly Thr Val Pro Val Arg Ala Leu Ala Gln Pro Gly Tyr  
 3010 3015 3020  
 Glu Ala Gly Glu Pro Val Pro Ala Ser Met Glu Ala Val Leu Gly Val  
 3025 3030 3035 3040  
 Gln Ala Asp Ala Val Leu Ala Ala Gln Gly Asp Thr Pro Phe Val Leu  
 3045 3050 3055  
 Val Gly His Ser Ala Gly Ala Leu Met Ala Tyr Ala Leu Ala Thr Glu  
 3060 3065 3070  
 Leu Ala Asp Arg Gly His Pro Pro Arg Gly Val Val Leu Leu Asp Val  
 3075 3080 3085  
 Tyr Pro Pro Gly His Gln Glu Ala Val His Ala Trp Leu Gly Glu Leu  
 3090 3095 3100  
 Thr Ala Ala Leu Phe Asp His Glu Thr Val Arg Met Asp Asp Thr Arg  
 3105 3110 3115 3120  
 Leu Thr Ala Leu Gly Ala Tyr Asp Arg Leu Thr Gly Arg Trp Arg Pro



```
<210> 16
<211> 358
<212> PRT
<213> Micromonospora megalomicea
```

48

340  
His Cys Pro Val Glu Leu  
355

345

350

<210> 17  
<211> 422  
<212> PRT  
<213> Micromonospora megalomicea

<400> 17  
Met Arg Val Val Phe Ser Ser Met Ala Ser Lys Ser His Leu Phe Gly  
1 5 10 15  
Leu Val Pro Leu Ala Trp Ala Phe Arg Ala Ala Gly His Glu Val Arg  
20 25 30  
Val Val Ala Ser Pro Ala Leu Thr Asp Asp Ile Thr Ala Ala Gly Leu  
35 40 45  
Thr Ala Val Pro Val Gly Thr Asp Val Asp Leu Val Asp Phe Met Thr  
50 55 60  
His Ala Gly Tyr Asp Ile Ile Asp Tyr Val Arg Ser Leu Asp Phe Ser  
65 70 75 80  
Glu Arg Asp Pro Ala Thr Ser Thr Trp Asp His Leu Leu Gly Met Gln  
85 90 95  
Thr Val Leu Thr Pro Thr Phe Tyr Ala Leu Met Ser Pro Asp Ser Leu  
100 105 110  
Val Glu Gly Met Ile Ser Phe Cys Arg Ser Trp Arg Pro Asp Trp Ser  
115 120 125  
Ser Gly Pro Gln Thr Phe Ala Ala Ser Ile Ala Ala Thr Val Thr Gly  
130 135 140  
Val Ala His Ala Arg Leu Leu Trp Gly Pro Asp Ile Thr Val Arg Ala  
145 150 155 160  
Arg Gln Lys Phe Leu Gly Leu Leu Pro Gly Gln Pro Ala Ala His Arg  
165 170 175  
Glu Asp Pro Leu Ala Glu Trp Leu Thr Trp Ser Val Glu Arg Phe Gly  
180 185 190  
Gly Arg Val Pro Gln Asp Val Glu Glu Leu Val Val Gly Gln Trp Thr  
195 200 205  
Ile Asp Pro Ala Pro Val Gly Met Arg Leu Asp Thr Gly Leu Arg Thr  
210 215 220  
Val Gly Met Arg Tyr Val Asp Tyr Asn Gly Pro Ser Val Val Pro Asp  
225 230 235 240  
Trp Leu His Asp Glu Pro Thr Arg Arg Arg Val Cys Leu Thr Leu Gly  
245 250 255  
Ile Ser Ser Arg Glu Asn Ser Ile Gly Gln Val Ser Val Asp Asp Leu  
260 265 270  
Leu Gly Ala Leu Gly Asp Val Asp Ala Glu Ile Ile Ala Thr Val Asp  
275 280 285  
Glu Gln Gln Leu Glu Gly Val Ala His Val Pro Ala Asn Ile Arg Thr  
290 295 300  
Val Gly Phe Val Pro Met His Ala Leu Leu Pro Thr Cys Ala Ala Thr  
305 310 315 320  
Val His His Gly Gly Pro Gly Ser Trp His Thr Ala Ala Ile His Gly  
325 330 335  
Val Pro Gln Val Ile Leu Pro Asp Gly Trp Asp Thr Gly Val Arg Ala  
340 345 350  
Gln Arg Thr Glu Asp Gln Gly Ala Gly Ile Ala Leu Pro Val Pro Glu  
355 360 365  
Leu Thr Ser Asp Gln Leu Arg Glu Ala Val Arg Arg Val Leu Asp Asp  
370 375 380  
Pro Ala Phe Thr Ala Gly Ala Ala Arg Met Arg Ala Asp Met Leu Ala  
385 390 395 400  
Glu Pro Ser Pro Ala Glu Val Val Asp Val Cys Ala Gly Leu Val Gly

405  
 Glu Arg Thr Ala Val Gly  
 420

410  
 415

<210> 18  
 <211> 323  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 18  
 Met Ser Thr Asp Ala Thr His Val Arg Leu Gly Arg Cys Ala Leu Leu  
 1 5 10 15  
 Thr Ser Arg Leu Trp Leu Gly Thr Ala Ala Leu Ala Gly Gln Asp Asp  
 20 25 30  
 Ala Asp Ala Val Arg Leu Leu Asp His Ala Arg Ser Arg Gly Val Asn  
 35 40 45  
 Cys Leu Asp Thr Ala Asp Asp Ser Ala Ser Thr Ser Ala Gln Val  
 50 55 60  
 Ala Glu Glu Ser Val Gly Arg Trp Leu Ala Gly Asp Thr Gly Arg Arg  
 65 70 75 80  
 Glu Glu Thr Val Leu Ser Val Thr Val Gly Val Pro Pro Gly Gly Gln  
 85 90 95  
 Val Gly Gly Gly Gly Leu Ser Ala Arg Gln Ile Ile Ala Ser Cys Glu  
 100 105 110  
 Gly Ser Leu Arg Arg Leu Gly Val Asp His Val Asp Val Leu His Leu  
 115 120 125  
 Pro Arg Val Asp Arg Val Glu Pro Trp Asp Glu Val Trp Gln Ala Val  
 130 135 140  
 Asp Ala Leu Val Ala Ala Gly Lys Val Cys Tyr Val Gly Ser Ser Gly  
 145 150 155 160  
 Phe Pro Gly Trp His Ile Val Ala Ala Gln Glu His Ala Val Arg Arg  
 165 170 175  
 His Arg Leu Gly Leu Val Ser His Gln Cys Arg Tyr Asp Leu Thr Ser  
 180 185 190  
 Arg His Pro Glu Leu Glu Val Leu Pro Ala Ala Gln Ala Tyr Gly Leu  
 195 200 205  
 Gly Val Phe Ala Arg Pro Thr Arg Leu Gly Gly Leu Leu Gly Gly Asp  
 210 215 220  
 Gly Pro Gly Ala Ala Ala Ala Arg Ala Ser Gly Gln Pro Thr Ala Leu  
 225 230 235 240  
 Arg Ser Ala Val Glu Ala Tyr Glu Val Phe Cys Arg Asp Leu Gly Glu  
 245 250 255  
 His Pro Ala Glu Val Ala Leu Ala Trp Val Leu Ser Arg Pro Gly Val  
 260 265 270  
 Ala Gly Ala Val Val Gly Ala Arg Thr Pro Gly Arg Leu Asp Ser Ala  
 275 280 285  
 Leu Arg Ala Cys Gly Val Ala Leu Gly Ala Thr Glu Leu Thr Ala Leu  
 290 295 300  
 Asp Gly Ile Phe Pro Gly Val Ala Ala Ala Gly Ala Ala Pro Glu Ala  
 305 310 315 320  
 Trp Leu Arg

<210> 19  
 <211> 247  
 <212> PRT  
 <213> Micromonospora megalomicea

<400> 19  
 Met Asn Thr Trp Leu Arg Arg Phe Gly Ser Ala Asp Gly His Arg Ala  
 1 5 10 15

Arg Leu Tyr Cys Phe Pro His Ala Gly Ala Ala Ala Asp Ser Tyr Leu  
 20 25 30  
 Asp Leu Ala Arg Ala Leu Ala Pro Glu Val Asp Val Trp Ala Val Gln  
 35 40 45  
 Tyr Pro Gly Arg Gln Asp Arg Arg Asp Glu Arg Ala Leu Gly Thr Ala  
 50 55 60  
 Gly Glu Ile Ala Asp Glu Val Ala Ala Val Leu Arg Asp Leu Val Gly  
 65 70 75 80  
 Glu Val Pro Phe Ala Leu Phe Gly His Ser Met Gly Ala Leu Val Ala  
 85 90 95  
 Tyr Glu Thr Ala Arg Arg Leu Glu Ala Arg Pro Gly Val Arg Pro Leu  
 100 105 110  
 Arg Leu Phe Val Ser Gly Gln Thr Ala Pro Arg Val His Glu Arg Arg  
 115 120 125  
 Thr Asp Leu Pro Asp Glu Asp Gly Leu Val Glu Gln Met Arg Arg Leu  
 130 135 140  
 Gly Val Ser Glu Ala Ala Leu Ala Asp Gln Gly Leu Leu Asp Met Ser  
 145 150 155 160  
 Leu Pro Val Leu Arg Ala Asp His Arg Val Leu Arg Ser Tyr Ala Trp  
 165 170 175  
 Gln Ala Gly Pro Pro Leu Arg Ala Gly Ile Thr Thr Leu Cys Gly Asp  
 180 185 190  
 Thr Asp Pro Leu Thr Thr Val Glu Asp Ala Gln Arg Trp Leu Pro Tyr  
 195 200 205  
 Ser Val Val Pro Gly Arg Thr Arg Thr Phe Pro Gly Gly His Phe Tyr  
 210 215 220  
 Leu Ala Asp His Val Gly Glu Val Ala Glu Ser Val Ala Pro Asp Leu  
 225 230 235 240  
 Leu Arg Leu Thr Pro Thr Gly  
 245

&lt;210&gt; 20

&lt;211&gt; 189

&lt;212&gt; PRT

&lt;213&gt; Micromonospora megalomicea

&lt;400&gt; 20

Ile Arg Val Gln Asp Asp Asp Ala Asp Arg Leu Ser Arg Asp Glu Leu  
 1 5 10 15  
 Thr Ser Ile Ala Leu Val Leu Leu Leu Ala Gly Phe Glu Ala Ser Val  
 20 25 30  
 Ser Leu Ile Gly Ile Gly Thr Tyr Leu Leu Leu Thr His Pro Asp Gln  
 35 40 45  
 Leu Ala Leu Val Arg Lys Asp Pro Ala Leu Leu Pro Gly Ala Val Glu  
 50 55 60  
 Glu Ile Leu Arg Tyr Gln Ala Pro Pro Glu Thr Thr Arg Phe Ala  
 65 70 75 80  
 Thr Ala Glu Val Glu Ile Gly Gly Val Thr Ile Pro Ala Tyr Ser Thr  
 85 90 95  
 Val Leu Ile Ala Asn Gly Ala Ala Asn Arg Asp Pro Gly Gln Phe Pro  
 100 105 110  
 Asp Pro Asp Arg Phe Asp Val Thr Arg Asp Ser Arg Gly His Leu Thr  
 115 120 125  
 Phe Gly His Gly Ile His Tyr Cys Met Gly Arg Pro Leu Ala Lys Leu  
 130 135 140  
 Glu Gly Glu Val Ala Leu Gly Ala Leu Phe Asp Arg Phe Pro Lys Leu  
 145 150 155 160  
 Ser Leu Gly Phe Pro Ser Asp Glu Val Val Trp Arg Arg Ser Leu Leu  
 165 170 175  
 Leu Arg Gly Ile Asp His Leu Pro Val Arg Pro Asn Gly  
 180 185

<210> 21  
 <211> 33  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Synthetic nucleotide DNA duplex

<400> 21  
 taagaattcg gagatctggc ctcagctcta gac 33

<210> 22  
 <211> 39  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> Complementary oligo

<400> 22  
 aattgtctag agctgaggcc agatctccga attcttaat 39

<210> 23  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<400> 23  
 ttgcagcggg tgctcgggtggc ggtgcggggag gggcgctcggg tgttggggtgt ggtgggtgggt 60  
 tcggcgggtga atcaggatgg ggcgagtaat ggggttggcgg cgccgctcggg ggtggcgag 120  
 cagcgggtga ttccggcgggc gtgggggtcgt gcgggtgtgt cgggtgggga tgtgggtgtg 180  
 gtggaggcgc atgggacggg gacgcgggtt ggggatccgg tggagttggg ggcgttggtg 240  
 gggacgtatg ggggtgggtcg ggggtggggtg ggtccggtgg tgggtgggttc ggtgaaggcg 300  
 aatgtgggtc atgtgcaggc ggcggcgggt gtgggtgggtg tgatcaagg ggtggtgggg 360  
 ttgggtcggg ggttgggtggg tccgatgggtg tgctcggggtg ggttgtcggg gttggtggat 420  
 tggtcgtcgg gtgggttggg ggtggcggat ggggtgcggg ggtggccggg ggggtgtggat 480  
 ggggtgcgtc ggggtggggg gtcggcggtt ggggtgtcgg ggacgaat 528

<210> 24  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<400> 24  
 ctgcagcggg tgctcgggtggc ggtgcggggag gggcgctcggg tgttggggtgt ggtgggtgggt 60  
 tcggcgggtga atcaggatgg ggcgagtaat ggggttggcgg cgccgctcggg ggtggcgag 120  
 cagcgggtga ttccggcgggc gtgggggtcgt gcgggtgtgt cgggtgggga tgtgggtgtg 180  
 gtggaggcgc atgggacggg gacgcgggtt ggggatccgg tggagttggg ggcgttggtg 240  
 gggacgtatg ggggtgggtcg ggggtggggtg ggtccggtgg tgggtgggttc ggtgaaggcg 300  
 aatgtgggtc atgtgcaggc ggcggcgggt gtgggtgggtg tgatcaagg ggtggtgggg 360  
 ttgggtcggg ggttgggtggg tccgatgggtg tgctcggggtg ggttgtcggg gttggtggat 420  
 tggtcgtcgg gtgggttggg ggtggcggat ggggtgcggg ggtggccggg ggggtgtggat 480  
 ggggtgcgtc ggggtggggg gtcggcggtt ggggtgtcgg ggacgaat 528

<210> 25  
 <211> 528  
 <212> DNA  
 <213> Micromonospora megalomicea

<220>

<221> misc\_feature

<222> (1)...(528)

<223> Sequence with codon changes as described in the  
specification at page 99, line 22 thru 101, line 23

<400> 25

ctgcagcgcc	tctccgtcgc	cgtccgcgag	ggccgcgcgag	tcctcggcgt	cgtcgtcggc	60
tcggccgtca	accaagacgg	cgcgtcaaac	ggcctcgccg	cgccctccgg	cgtcgcccag	120
cagcgcgta	tacgccgcgc	gtggggacgc	gccggagtat	cgggcggcga	cgtcggagtc	180
gtcagggccc	acggcaccgg	cacccgcctc	ggggatcccc	tcgagctggg	cgccctcctg	240
ggcacgtacg	gcgtcggccg	cggcggcgtc	ggcccggtcg	tcgtcggcag	cgtcaaggcc	300
aacgtcggcc	acgtccaggc	cgcggccggc	gtcgtcgggg	tcacaaaggt	cgtcctcggc	360
ctcggccgcg	ggctggtcgg	cccgatggtc	tgccgcggcg	gcctcagcgg	cctcgtcgcg	420
tggtcgtccg	gcggcctggg	cgtcgcggac	ggggtcgcgc	gctggccggg	cggcgtcgcg	480
ggcgtccgcc	ggggcggcgt	ctcggcgctt	ggcgtcagcg	ggacgaat		528

<210> 26

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<400> 26

ggtggagtgt	gatgcggtgg	tgctcgtcgt	ggtgggggtt	tcgggtgttg	gggtgttgga	60
gggtcggtcg	ggtgcgccgt	cggttgatcg	ggtggatgtg	gtgcagccgg	tggtgttcgt	120
ggtgatggtg	tcgttggcgc	ggttggtggc	gtgggtgtgg	ggttggtcctg	cggcgggtgg	180
gggtcattcg	cagggggaga	tcgcggcgcc	ggtgggtggc	gggggtgttg	cgggtgggtga	240
tggtgcgcgg	gtgggtggcg	tgccggcgcg	ggcgttgccg	gcgttgggcg	g	291

<210> 27

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<400> 27

ggtggagtgt	gatgcggtgg	tgctcgtcgt	ggtgggggtt	tcgggtgttg	gggtgttgga	60
gggtcggtcg	ggtgcgccgt	cggttgatcg	ggtggatgtg	gtgcagccgg	tggtgttcgt	120
ggtgatggtg	tcgttggcgc	ggttggtggc	gtgggtgtgg	ggttggtcctg	cggcgggtgg	180
gggtcattcg	cagggggaga	tcgcggcgcc	ggtgggtggc	gggggtgttg	cgggtgggtga	240
tggtgcgcgg	gtgggtggcg	tgccggcgcg	ggcgttgccg	gcgttgggcg	g	291

<210> 28

<211> 291

<212> DNA

<213> Micromonospora megalomicea

<220>

<221> misc\_feature

<222> (1)...(291)

<223> Sequence with codon changes as described in the  
specification at page 99, line 22 thru page 101, line 23

<400> 28

cgtggagtgc	gatgcggtcg	tgctcagcgt	cgtcggcttc	agcgtgctgg	gcgtcctgga	60
gggcccgcgc	ggcgcgccga	gcctggaccg	cgtcgcagtg	gtccagccgg	tcctgttcgt	120
ggtcatggtc	agcctggccc	gcctgtggcg	ctgggtgcgc	gtgggtcccg	ccgccgtggg	180
cggccacagc	cagggcgaga	tcgccgcgcg	ggctcgtggc	ggcgtcctga	gcgtcggcga	240
cggcgcgccg	gtcgtggccc	tgccgcgccg	cgccctgcgc	gccctggccg	g	291

<210> 29

<211> 24

<212> DNA



<213> Artificial Sequence

<220>

<223> PCR primer

<400> 29

gaacaactcc tgtctgcggc cgcg

24

<210> 30

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 30

cggaattctc tagagtcacg tctccaaccg cttgtcgagg

40

<210> 31

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 31

tctagactta attaaggagg acacatatga gcgagagcag cggcatgacc g

51

<210> 32

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 32

aacgcctccc aggagatctc cagca

25

<210> 33

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligo

<400> 33

aattcatagc ctaggt

16

<210> 34

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligo

<400> 34

# INTERNATIONAL SEARCH REPORT

International Application No  
PC1/US 00/27433

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/52 C12N15/53 C12N15/54 C12N15/61 C12N15/62 C12N9/04 C12N9/10 C12N9/90 C12P19/62		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C12P		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EMBL, EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, EMBASE		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 23630 A (ABBOTT LAB) 3 July 1997 (1997-07-03) the whole document claims 1-22 figures 1-3	1-12, 14, 18, 19
X	WO 99 05283 A (MENDEZ CARMEN ; SALAS JOSE A (ES); RAYNAL MARIE CECILE (FR); FROMEN) 4 February 1999 (1999-02-04) the whole document claims 1-41	1-12, 14, 18, 19
--- -/--		
<div style="display: flex; justify-content: space-between;"> <span><input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.</span> <span><input checked="" type="checkbox"/> Patent family members are listed in annex.</span> </div>		
<div style="display: flex;"> <div style="flex: 1;"> <p>* Special categories of cited documents:</p> <p>*A* document defining the general state of the art which is not considered to be of particular relevance</p> <p>*E* earlier document but published on or after the international filing date</p> <p>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>*O* document referring to an oral disclosure, use, exhibition or other means</p> <p>*P* document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="flex: 1;"> <p>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>*8* document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search  <div style="text-align: center; font-weight: bold;">13 June 2001</div>		Date of mailing of the international search report  <div style="text-align: center; font-weight: bold;">09/07/2001</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer  <div style="text-align: center; font-weight: bold;">van de Kamp, M</div>

# INTERNATIONAL SEARCH REPORT

International Application No  
PCI/US 00/27433

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SUMMERS R G ET AL.: "Sequencing and mutagenesis of genes from the erythromycin biosynthetic gene cluster of <i>Saccharopolyspora erythraea</i> that are involved in L-mycarose and D-desosamine production" MICROBIOLOGY, vol. 143, 1 October 1997 (1997-10-01), pages 3251-3262, XP002061260 cited in the application abstract page 3253, right-hand column, line 47 -page 3253, left-hand column, line 19 figures 1-6; table 1	1-12,14, 18,19
X	OLANO C ET AL.: "Analysis of a <i>Streptomyces antibioticus</i> chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring" MOLECULAR AND GENERAL GENETICS, vol. 259, no. 3, 1 August 1998 (1998-08-01), pages 299-308, XP002096258 cited in the application abstract page 300, right-hand column, line 46 -page 301, left-hand column, line 17 figures 1,2	1,5-12, 19
X	XUE Y ET AL.: "A gene cluster for macrolide antibiotic biosynthesis in <i>Streptomyces venezuelae</i> : architecture of metabolic diversity" PROC. NATL. ACAD. SCI. USA, vol. 95, October 1998 (1998-10), pages 12111-12116, XP002166926 cited in the application abstract page 12113, left-hand column, line 4-24 figures 1,2; tables 1,2	1,5-12, 19
X	OTTEN S L ET AL.: "Cloning and chracterization of the <i>Streptomyces peucetius</i> dmZUV genes encoding three enzymes required for biosynthesis of the daunorubicin precursor thymidine diphospho-L-daunosamine" JOURNAL OF BACTERIOLOGY, vol. 179, no. 13, July 1997 (1997-07), pages 4446-4450, XP002166927 abstract figure 1; table 1	1,5-12, 19

-/--

# INTERNATIONAL SEARCH REPORT

International Application No

PC1/US 00/27433

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>OTTEN S L ET AL.: "Cloning and characterization of the Streptomyces peucetius dnrQS genes encoding a daunosamine biosynthesis enzyme and a glycosyl transferase involved in daunorubicin biosynthesis" JOURNAL OF BACTERIOLOGY, vol. 177, no. 22, November 1995 (1995-11), pages 6688-6692, XP002166928 abstract figure 1</p>	1,5-12, 19
X	<p>TORKKELL S ET AL.: "Characterization of Streptomyces nogalater genes encoding enzymes involved in glycosylation steps in nogalamycin biosynthesis" MOLECULAR AND GENERAL GENETICS, vol. 256, no. 2, September 1997 (1997-09), pages 203-209, XP002166929 cited in the application abstract figure 1</p>	1,5-12, 19
A	<p>SWAN D G ET AL.: "Characterisation of a Streptomyces antibioticus gene encoding a type I polyketide synthase which has an unusual coding sequence" MOLECULAR AND GENERAL GENETICS, vol. 242, no. 3, 1994, pages 358-362, XP002087278 cited in the application abstract page 358, right-hand column, line 5 -page 361, left-hand column, line 18</p>	1,9
Y	<p>US 3 819 611 A (WEINSTEIN M ET AL) 25 June 1974 (1974-06-25) the whole document</p>	1-12,14, 18-20
Y	<p>MALPARTIDA F ET AL: "Homology between Streptomyces genes coding for synthesis of different polyketides used to clone antibiotic biosynthetic genes" NATURE, vol. 325, 26 February 1987 (1987-02-26), pages 818-821, XP002075972 abstract</p>	1-12,14, 18-20
A	<p>NAKAGAWA A ET AL.: "Structure and stereochemistry of macrolides" MACROLIDE ANTIBIOTICS. OMURA S (ED.). PUBLISHER: ACADEMIC, ORLANDO, FLORIDA, 1984, pages 37-84, XP001006199 page 46, line 25 -page 48, line 4</p>	

-/--

## INTERNATIONAL SEARCH REPORT

International Application No

PCI/US 00/27433

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LEONARD KATZ: "Manipulation of modular polyketide synthases" CHEMICAL REVIEWS, vol. 97, no. 7, 1997, pages 2557-2575, XP002103748 the whole document	1-12,14, 18,19
A	LIU H -W ET AL: "Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria" ANNUAL REVIEW OF MICROBIOLOGY, vol. 48, 1994, pages 223-256, XP002061259 page 234, line 24 -page 237, line 9; figures 8,9	1,5-12, 19
A	CARRERAS C W ET AL.: "Engineering of modular polyketide synthases to produce novel polyketides" CURRENT OPINION IN BIOTECHNOLOGY, vol. 9, no. 4, August 1998 (1998-08), pages 403-411, XP000993508 the whole document	14,18
A	HUTCHINSON C R: "Combinatorial biosynthesis for new drug discovery" CURRENT OPINION IN MICROBIOLOGY, vol. 1, no. 3, June 1998 (1998-06), pages 319-329, XP000993550 the whole document	14,18
A	MCDANIEL R ET AL.: "Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel unnatural natural products" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 96, 1999, pages 1846-1851, XP000910246 cited in the application abstract	14,18
P,X	VOLCHEGURSKY Y ET AL.: "Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in Saccharopolyspora erythrea" MOLECULAR MICROBIOLOGY, vol. 37, no. 4, August 2000 (2000-08), pages 752-762, XP002166930 the whole document	1-6, 8-13, 18-20
P,X	WO 00 00500 A (LEADLAY PETER FRANCIS ;CORTES JESUS (GB); STAUNTON JAMES (GB); BIO) 6 January 2000 (2000-01-06) claim 24	14

-/--

# INTERNATIONAL SEARCH REPORT

International Application No  
PC/US 00/27433

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	<p>WO 00 63361 A (KOSAN BIOSCIENCES INC)  26 October 2000 (2000-10-26)  page 9, line 3-9  page 14, line 26 -page 16, line 2  claim 3</p>	<p>1-13,  18-20</p>



# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCI/US 00/27433

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9723630 A	03-07-1997	US 5998194 A EP 0874548 A JP 2000502899 T	07-12-1999 04-11-1998 14-03-2000
WO 9905283 A	04-02-1999	FR 2766496 A FR 2786200 A EP 1032679 A	29-01-1999 26-05-2000 06-09-2000
US 3819611 A	25-06-1974	BE 715638 A CA 931891 A CH 534206 A CS 157635 B DE 1767565 A DK 123422 B ES 354296 A FI 46519 B FR 8066 M GB 1229835 A IE 31918 B IL 30067 A LU 56131 A NL 6807363 A NO 128225 B OA 4027 A SE 349323 B	25-11-1968 14-08-1973 28-02-1973 16-09-1974 14-10-1971 19-06-1972 16-10-1969 02-01-1973 06-07-1970 28-04-1971 07-02-1973 28-09-1972 11-09-1968 27-11-1968 15-10-1973 15-09-1979 25-09-1972
WO 0000500 A	06-01-2000	AU 4524599 A AU 4524799 A BR 9911710 A BR 9911712 A EP 1091971 A EP 1090123 A WO 0000618 A	17-01-2000 17-01-2000 20-03-2001 20-03-2001 18-04-2001 11-04-2001 06-01-2000
WO 0063361 A	26-10-2000	AU 4241800 A	02-11-2000